AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology and Molecular Biology Reviews

Regulatory RNAs in *Bacillus subtilis*: a Gram-Positive Perspective on Bacterial RNA-Mediated Regulation of Gene Expression

Ruben A. T. Mars,[a]* Pierre Nicolas,[b] Emma L. Denham,[a]* Jan Maarten van Dijl[a]

Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands[a]; MaIAGE, INRA, Université Paris-Saclay, Jouy-en-Josas, France[b]

SUMMARY

Bacteria can employ widely diverse RNA molecules to regulate their gene expression. Such molecules include *trans*-acting small regulatory RNAs, antisense RNAs, and a variety of transcriptional attenuation mechanisms in the 5′ untranslated region. Thus far, most regulatory RNA research has focused on Gram-negative bacteria, such as *Escherichia coli* and *Salmonella*. Hence, there is uncertainty about whether the resulting insights can be extrapolated directly to other bacteria, such as the Gram-positive soil bacterium *Bacillus subtilis*. A recent study identified 1,583 putative regulatory RNAs in *B. subtilis*, whose expression was assessed across 104 conditions. Here, we review the current understanding of RNA-based regulation in *B. subtilis*, and we categorize the newly identified putative regulatory RNAs on the basis of their conservation in other bacilli and the stability of their predicted secondary structures. Our present evaluation of the publicly available data indicates that RNA-mediated gene regulation in *B. subtilis* mostly involves elements at the 5′ ends of mRNA molecules. These can include 5′ secondary structure elements and metabolite-, tRNA-, or protein-binding sites. Importantly, sense-independent segments are identified as the most conserved and structured potential regulatory RNAs in *B. subtilis*. Altogether, the present survey provides many leads for the identification of new regulatory RNA functions in *B. subtilis*.

INTRODUCTION

While RNA was initially seen as merely the messenger between genetic information present in the DNA and the cellular machinery composed of proteins, the importance of posttranscriptional regulation by regulatory RNA molecules is now widely appreciated in virtually all studied organisms (1). Regulatory RNAs were first identified in bacteria in the 1980s, years before the first microRNAs (miRNAs), short interfering RNAs (siRNAs), and long noncoding RNAs (lncRNAs) in eukaryotes were reported (1–3). However, it was not until the beginning of the 2000s that knowledge on bacterial regulatory RNAs started to expand

Published 26 October 2016

Citation Mars RAT, Nicolas P, Denham EL, van Dijl JM. 2016. Regulatory RNAs in *Bacillus subtilis*: a Gram-positive perspective on bacterial RNA-mediated regulation of gene expression. Microbiol Mol Biol Rev 80:1029–1057. doi:10.1128/MMBR.00026-16.

Address correspondence to Jan Maarten van Dijl, j.m.van.dijl01@umcg.nl.

* Present address: Ruben A. T. Mars, Institute of Molecular Systems Biology, ETH Zürich, Zürich, Switzerland; Emma L. Denham, Division of Biomedical Sciences, Warwick Medical School, University of Warwick, Coventry, United Kingdom.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/MMBR.00026-16.

dramatically (2). Bacterial regulatory RNAs are now recognized as a heterogeneous group of molecules that act by a wide variety of mechanisms to modulate a plethora of physiological responses.

Many RNA molecules carry out housekeeping functions in a variety of ways. They can target incomplete proteins for degradation and release stalled ribosomes (transfer-messenger RNA [tmRNA]), form a structural component of the signal recognition particle (4.5S RNA) that modulates protein export from the cytoplasm, present amino acids to the ribosome (tRNA), and process tRNA or mRNA (RNase P RNA) (2). Another class of regulatory RNAs is cotranscribed with protein-encoding mRNA. These *cis*-acting RNA elements can be attached to the 5′ or 3′ ends of mRNA molecules or situated between open reading frames (ORFs) on multicistronic mRNA to enable control over RNA degradation, RNA processing, or ribosome progression. Another comprehensively studied class of RNA regulators is the *trans*-acting small regulatory RNAs, which act by short, imperfect, complementary base pairing to specific mRNAs to thereby modulate the mRNA stability and/or extent of translation. Last, a special group of RNA regulators is known as CRISPR RNAs (clustered regularly interspaced short palindromic repeat RNAs), which act as a bacterial memory against bacteriophage and plasmid sequences, analogously to an innate immune system (4).

RNA molecules have an astounding structural and functional flexibility. This flexibility is exploited in their regulatory mechanisms, which often entail conformational changes upon an environmental trigger. These triggers may, for instance, be the binding of metabolites, binding of proteins, or changes in temperature. A frequently occurring outcome of such binding events is the modulation of translation efficiency, which can take place through the occlusion or exposure of a ribosome-binding site upon small RNA (sRNA)-mediated remodeling of the local mRNA structure. Despite the complications of RNA secondary structure predictions, the energetically most stable RNA structure can be computed quite accurately *in silico* (5). The predicted base pairs in such structures are, however, not perfect and must be interpreted with caution. For sequences of <700 bases, ~70% of the base pairs can be expected to be predicted correctly, but this drops to 20 to 60% for longer sequences (6). In addition, the predicted structure is not necessarily the most dominant and functional form *in vivo*. Nevertheless, the shortcomings of *in silico* RNA folding analyses do not render them meaningless. For example, *in silico* RNA folding studies revealed that predicted exposed bases are more likely to participate in sRNA-mRNA interactions (7), help in predicting sRNA targets (8), and allow a clear distinction between known regulatory RNAs and random sequences (9).

Regulatory RNAs have been identified by a variety of experimental approaches. Initially, these involved systematic bioinformatic searches for conserved intergenic sequence elements with accompanying orphan promoters and terminators in *Escherichia coli* (10). This was followed by the development of microarrays specific to intergenic regions (2, 11, 12). Later, more sophisticated approaches were developed, including advanced computational searches, transcript profiling with high-density tiling arrays, and RNA sequencing (RNA-seq). These studies have greatly enhanced our understanding of the complexity of bacterial transcriptomes (13–17). A subset of these studies explicitly focused on annotating all transcriptionally active regions, and this enabled the discovery of large numbers of putative regulatory RNAs (13, 15).

One large-scale tiling array study was conducted with the *Bacillus subtilis* strain 168 and addressed genome-wide transcription under 104 different environmental conditions. In this study, ~85% of all genes were found to be highly expressed under at least one condition, 4.4% of the genes were not expressed under any condition, and ~3% of the genes were highly expressed under all conditions (15). This implies that the applied conditions covered the largest part of the transcriptional space of *B. subtilis*. Among the goals of this study were the accurate mapping of promoters and new RNA segments. Despite the strict expression cutoffs that were applied, the study reported almost 1,600 putative regulatory RNAs that had previously remained unannotated in the *B. subtilis* GenBank records. These 1,583 identified RNA segments were annotated with an S number, where the "S" stands for segment, and their expression profiles can be explored at http://genome.jouy.inra.fr/cgi-bin/seb/index.py. It should be noted that functions for approximately 60 of these S segments had been reported before but that, at the time, these segments were not included in the employed GenBank annotation (see Table S1 in the supplemental material). Altogether, the set of 1,583 putative regulatory RNAs represents a valuable resource for expanding our understanding of RNA regulation in Gram-positive bacteria in general and *B. subtilis* in particular, especially since the diversity of the employed expression conditions allows for the design of specific follow-up studies.

SCOPE OF THE REVIEW

In the present review, we aim at providing a compendium and critical evaluation of the previously identified RNA segments of *B. subtilis*. This is relevant from both fundamental scientific and biotechnological perspectives. For many years, the *Firmicutes* species *B. subtilis* has been one of the key paradigms for studies on prokaryotic cell division as well as cell differentiation, with special focuses on processes such as motility; the development of genetic competence for external DNA binding, uptake, and recombination; and the formation of heat-resistant endospores (18, 19). Additionally, *B. subtilis* has become one of the major workhorses in biotechnology, where it is mostly applied in the commercial production of secreted technical enzymes and vitamins (20, 21). Interestingly, in both fundamental and applied investigations of *B. subtilis*, the level of RNA-based regulation has received relatively little attention. This might be linked to the reported minor importance of the RNA chaperone Hfq to RNA regulation in most Gram-positive bacteria, and *B. subtilis* in particular. In contrast, the critical role of Hfq in Gram-negative bacteria greatly facilitated the many studies on RNA regulation in the respective species (22–24). To bridge current knowledge gaps, we now provide a global description of all identified regulatory RNAs of *B. subtilis*. This covers not only regulatory RNAs with assigned physiological functions but also all newly identified putative regulatory RNAs. Specifically, we detail the different groups of known regulatory elements, address the evolutionary conservation and *in silico* predicted secondary structures of newly identified segments, and discuss possible functions of the newly identified segments, as summarized in Fig. 1. In doing so, we identify novel opportunities for future detailed studies of RNA regulation in *B. subtilis*, analogously to previous reviews on signal peptides that proved instrumental for subsequent studies on protein secretion (25, 26). Notably, regulatory RNAs in other low-GC Gram-positive bacteria were fairly recently reviewed by Brantl and Brückner (27), and these RNAs are not addressed here unless it is relevant for a better

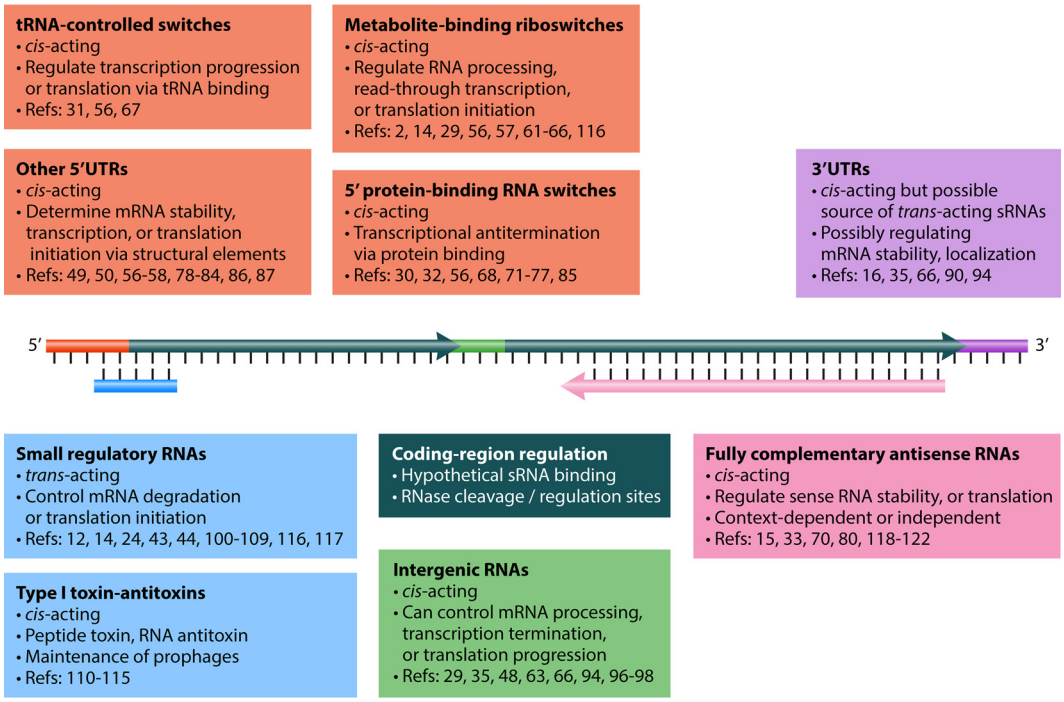

FIG 1 Graphic summary of possible regulatory RNA functions.

understanding of homologous or analogous regulatory RNAs of *B. subtilis*. Furthermore, this review does not cover 6S RNA and the various housekeeping RNAs mentioned in the Introduction, and it only briefly introduces riboswitches, T-box RNA switches, and protein-binding RNA switches with references to relevant publications or reviews (2, 28–32).

FUNCTIONAL REGULATORY RNAs AND NEWLY IDENTIFIED RNA SEGMENTS IN *B. SUBTILIS*

General Description of Newly Identified Putative Regulatory RNAs

The 1,583 newly identified putative regulatory RNAs (15) were initially divided into eight categories based on their locations with respect to the nearest protein coding DNA sequences (CDS), the consistency of their expression levels in the complete condition space, and the extents of transcriptional termination. In order to facilitate the descriptions in this review, we considered a simplified version of this classification consisting of four categories: 5′, all-3′, all-intergenic, and all-independent. The first category (5′) is formed by the 5′ untranslated regions (5′UTRs). The second category (all-3′) encompasses three types of 3′UTRs, the first with clear termination signals (3′), the second with partial termination (3′-PT), and the third without termination (3′-NT). It is currently unclear whether there is a biological meaning to the observed incomplete termination. However, it has been suggested that this phenomenon may be indicative of spuriously transcribed segments and may thus be the result of lowered purifying selection (15, 33). For the present review, these three types of 3′UTR segments were pooled into one category, referred to as all-3′. Besides the 5′UTR and 3′UTR segments, Nicolas et al. distinguished two types of intergenic regions, namely, intergenic regions that are situated between genes under the control of one promoter (Intra segments) and intergenic regions that are situated between genes under the control of distinct promoters (Inter segments). For this review, these two types of intergenic regions were also pooled into one category, referred to as all-intergenic (i.e., category 3). The final category of RNA segments is composed of the independently transcribed segments, referred to as Indep segments. These RNA segments contain their own transcription and termination elements and are therefore potentially related to the sRNAs. A subgroup of these Indep segments, referred to as Indep-NT segments, also have their own promoters but are not associated with a clear terminator. Because segments from the last two categories may be independent regulators, they were pooled into one category for the present analyses, which is referred to as all-independent (i.e., category 4).

Inherent to the analysis of RNA with tiling arrays is the uncertainty of transcriptional start and stop sites. This is a downside of tiling arrays compared to some RNA-seq protocols, where the RNA ends can be inferred directly from the sequencing data (34). In the study by Nicolas et al., the tiling step was 22 nucleotides (nt) on each strand of the genome (~400,000 probes in total) (15, 35), which resulted in transcript ends determined with a precision of around ±11 nucleotides (15, 35). This means that determination of RNA start and end sites may still be desired for more detailed studies. This could, for instance, be done by rapid amplification of the RNA ends or by dedicated RNA-seq protocols (13, 17, 36). Notably, some start and end sites were already known for the previously published regulatory RNAs or RNAs that were identified in an RNA-seq analysis by Irnov et al. (14).

Clues for Predicting Regulatory RNA Functions

Identifying a function for a regulatory RNA can be challenging and laborious. We distinguish five aspects of RNA gene organiza-

tion that can provide clues toward their functions and thus facilitate the design of biological experiments to verify the predicted functions. These five aspects are evolutionary conservation, genomic context, transcription factor control, expression level, and expression correlation.

The first aspect is the conservation of the RNA sequence of interest, its expression in other species, and, obviously, the identification of a function in other species where this RNA is conserved. The most convenient ways of finding conserved RNA sequences involve the use of databases, such as the RFAM database of conserved RNA families (37; http://rfam.xfam.org/), or the identification of sequence elements linked to a particular regulatory mechanism, for instance, with the online RibEx algorithm (38; http://132.248.32.45/cgi-bin/ribex.cgi). However, such databases include only known regulatory sequences, while it is anticipated that the majority of the diversity in RNA regulation has not yet been identified. In addition, we noted that many segments for which we observed significant RFAM motifs resulted from incomplete genome annotation. This means that the respective regulatory elements had been described but not yet annotated in the *B. subtilis* 168 genome sequence used for the tiling array study by Nicolas et al. (15, 39) (see Table S1 in the supplemental material).

The second aspect for the prediction of RNA function is the genomic region in which the RNA segment is located. For instance, genes encoding toxin-antitoxin systems are found mostly in prophage regions, which are phage genomes that have integrated into the bacterial chromosome. At a more local level, the functional annotation of genes in the direct vicinity of an RNA gene can give an indication of its function. This is especially the case when their promoters are situated on the same strand and show a correlation in their expression profile. Specifically, the annotation of downstream genes is a useful criterion for riboswitch annotation, as exemplified by genes involved in purine metabolism that have a high chance of being preceded by a purine riboswitch. In addition, an independently transcribed RNA segment between two important sporulation genes is likely to be involved in the process of spore formation.

The third aspect that may point to a function of an RNA segment is its control by a particular transcriptional regulator. Transcriptional regulator-binding motifs in well-studied model bacteria can be predicted with some certainty based on degenerate sequence motifs (40, 41). Such an identification of transcriptional regulators can start with a search for sigma factor-binding motifs, since some sigma factors are linked to specific processes, as is the case in *B. subtilis*, which employs a large number of alternative sigma factors for different processes (42). For example, a highly significant hit for a binding site of one of the sporulation-specific sigma factors will point to a role in sporulation, and analogously, a highly significant hit for a binding site of the stress sigma factor SigB points to a role in the SigB-controlled stress adaptation pathway. The presence of potential binding sites for other protein factors can also give clues to the regulatory process controlled by a particular RNA. For instance, for *B. subtilis*, this was reported for the Fur-dependent FsrA sRNA, which regulates the iron-sparing response (43), and the gluconeogenesis sRNA SR1, which is regulated by CcpA and CcpN (44).

Obviously linked to the association with potential regulators is the fourth aspect that can be addressed to predict RNA function, namely, the specific expression level of an RNA segment under a defined environmental condition. For example, when an independently expressed RNA is specifically induced under conditions of sporulation or stress, this provides a clue to a potential role in the respective process where the RNA is induced. The wealth of expression conditions addressed by Nicolas et al. (15) gives an unprecedented view of the condition-dependent expression of protein-encoding genes and potential regulatory RNA genes.

In turn connected to the respective expression level is the fifth and final aspect that can be applied to predict RNA function, namely, that of expression correlation. Expression correlation over a wide range of experimental conditions was applied by Nicolas et al. in promoter cluster analyses to unravel regulons, and accordingly, this information can be used to infer indications for regulatory RNA functions (15). This seems especially effective for predicting the functions of antisense RNAs (asRNAs). If an asRNA is, for instance, highly negatively correlated with its sense RNA, this can point to a specific regulatory function. The five aspects that can be applied to predict RNA function are discussed for individual cases of putative regulatory RNAs in the following sections, with focuses on the four different RNA categories (i.e., 5′, all-3′, all-intergenic, and all-independent) as defined above.

Sigma Factor Regulation, Structural Predictions, and Conservation Analysis of Newly Identified Putative Regulatory RNAs

Since some sigma factors are specifically induced under certain experimental conditions, the regulation of a regulatory RNA by a particular sigma factor can point to a function of this RNA. Sigma factor regulation of all promoter regions (i.e., transcriptional upshifts) in *B. subtilis* was predicted for all segments using an advanced sequence-based statistical model (15). From this analysis, we plotted the nearest promoter of all putative regulatory RNAs per category (Fig. 2). For all RNA categories, the household sigma factor SigA was found to be the most prominent sigma factor. Notably, SigA-dependent regulation does not necessarily imply constitutive expression, since these promoters are often under the control of additional factors (45). In this analysis, the all-independent RNA category was most different from the three other categories. This category contains a much larger fraction of segments regulated by alternative sigma factors, most prominently SigB, SigE, SigF, and SigK, than the other three categories. This may reflect a more prominent role for regulation by sRNAs under conditions of stress and sporulation. The predicted segment-specific sigma factor dependency is listed for all RNA segments in an overview table with all RNA segments (see Table S2 in the supplemental material).

Conservation throughout evolution is an important indication of the functional importance of a gene. To identify RNA segments that are likely biologically relevant, we analyzed the evolutionary conservation of all new putative regulatory RNAs in a set of 62 *Bacillus* genomes. Notably, we found that the asRNA segments display a much higher level of evolutionary conservation than sense RNA segments. This is a logical consequence of the fact that asRNAs are complementary to protein-encoding RNAs, which are generally more conserved than regulatory RNAs. This intrinsic problem makes the analysis of the evolutionary conservation of asRNAs challenging. However, it is possible to analyze the evolutionary conservation of an asRNA promoter, and this type of analysis was previously described for *B. subtilis* by Nicolas et al. (see SOM Fig. S23 in reference 15). To avoid a bias in the conservation analysis of RNA segments from *B. subtilis*, asRNAs were excluded

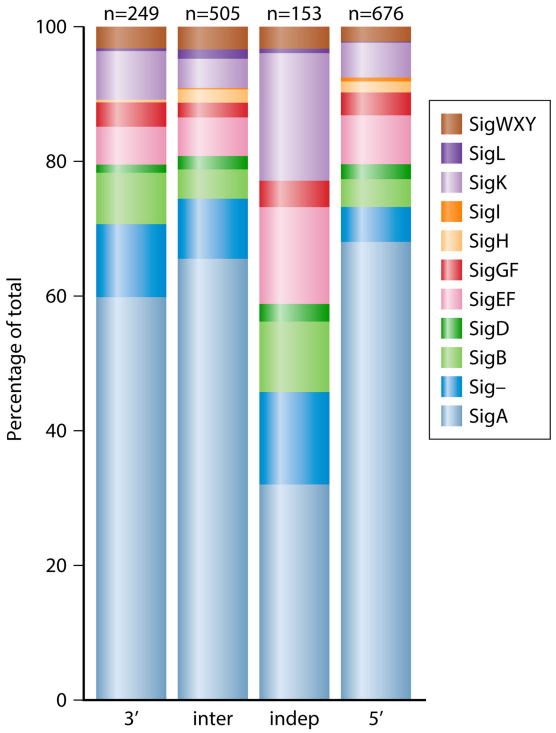

FIG 2 Sigma factor distribution per RNA category. The annotation from the nearest promoter was plotted. SigA is the most dominant sigma factor for all RNA categories, and the all-independent segments have the largest diversity in their sigma factor regulation. Colors correspond to the sigma factors and are aligned with respect to the column on the far right (5′UTR).

when plotting evolutionary conservation in the following sections of this paper.

To visualize the fraction of sense RNA segment conservation in each investigated genome, we plotted the evolutionary conservation of these segments in a heat map representation (Fig. 3). When inspecting this global conservation analysis, two points are notable. First, the considered genomes cluster in three groups based on the conservation of the putative regulatory RNA segments. The first group contains only genomes of *B. subtilis* isolates (including strain 168), and almost all investigated RNA segments are conserved in these genomes. The second group consists of genomes in which approximately half of the segments are conserved (consisting of more distantly related *B. subtilis* isolates, *Bacillus amyloliquefaciens*, and *Bacillus atrophaeus*). The third group of genomes displays a relatively low level of evolutionary conservation, and this is consistent with their greater evolutionary distance. The automated reordering of the genomes in rows based on RNA segment conservation therefore reflects the evolutionary relatedness of the species. Second, there are RNA segments that are highly conserved in almost all considered genomes, and these belong in particular to the classes all-independent and 5′. This higher conservation of the all-independent and 5′ RNA segments is also apparent and significant when sequence conservation (now expressed as the number of genomes with significant nBLAST hits) per group is inspected in more detail (Fig. 4A). This thus suggests a larger role for regulation via 5′UTRs and independent segments than via 3′UTRs and intergenic regions in *B. subtilis*. Yet, while a lack of evolutionary conservation may point to a lack of function, it should be kept in mind that some conserved RNAs may still lack a distinct biological function.

The structure and stability of an RNA molecule are important for its regulatory capabilities, and several algorithms have been developed to predict stable RNA structures (6). We decided to predict the secondary structures of all (new) RNA segments using RNAfold (46). RNAfold uses a loop-based energy model and scans the entire landscape of possible secondary structure configurations to identify the thermodynamically most stable structure. Because the minimum free energy (MFE) was found to be strongly sequence length dependent, we computed secondary structure MFE Z-scores (segment MFE − mean MFE of 100 shuffled sequences with the same length and nucleotide composition/standard deviation of the MFE between the 100 sequences) for all segments and used these as a measure of secondary structure for all investigated RNA segments.

While the level of predicted secondary structure within all putative regulatory RNA categories is considerable (Fig. 4B), the median MFE Z-score of the putative sRNA category (all-independent) is the lowest. This highly significant observation (Fig. 4B) means that, next to their relatively high evolutionary conservation, the all-independent segments also have the largest degree of secondary structure. Conversely, the 5′UTR category has the smallest degree of predicted secondary structure. The latter may indirectly be the result of terminator-associated stem-loop structures that can be present at the 3′ end of RNA molecules but not the 5′ end. Alternatively, the lower level of predicted secondary structure in the 5′UTRs could reflect the known lower secondary structure of the translation initiation region (47, 48).

We subsequently examined the RNA secondary structure in combination with the species-level conservation. Here, a small but significant correlation was found between stronger secondary structures (i.e., lower MFE Z-scores) and higher RNA segment conservation (Fig. 5A). This correlation was stronger when asRNA segments were excluded from the analysis (Fig. 5B), which probably reflects the generally lower degree of structure in asRNAs than in sense RNAs. Furthermore, this correlation is significant for the categories all-independent and all-intergenic but not for categories 5′ and all-3′ (Fig. 6). The observed correlation between stronger secondary structure and increased segment conservation seems to support the idea that functional, and thus more important, regulatory RNAs are more structured than nonfunctional RNA segments.

The following sections will review known *B. subtilis* regulatory RNA segments grouped by five categories: (i) RNA regulatory elements in the 5′ untranslated region, including metabolite-binding riboswitches, tRNA-controlled RNA switches, protein-binding RNA switches, and 5′UTRs with other or unknown functions; (ii) 3′UTRs; (iii) intergenic regions; (iv) independently expressed small RNAs and RNA antitoxins; and (v) antisense RNAs. Each section will also include a discussion of multiple segments that may be functional regulatory RNAs based on the bioinformatic evaluations provided with this review and data from the work of Nicolas et al. (15).

RNA Regulatory Elements in the 5′ Untranslated Region

5′UTRs are RNAs that are cotranscribed with protein-encoding mRNAs from a promoter further upstream of the translation start site. Their mode of regulation is directly tied to their adopted secondary structure, since this secondary structure can control the

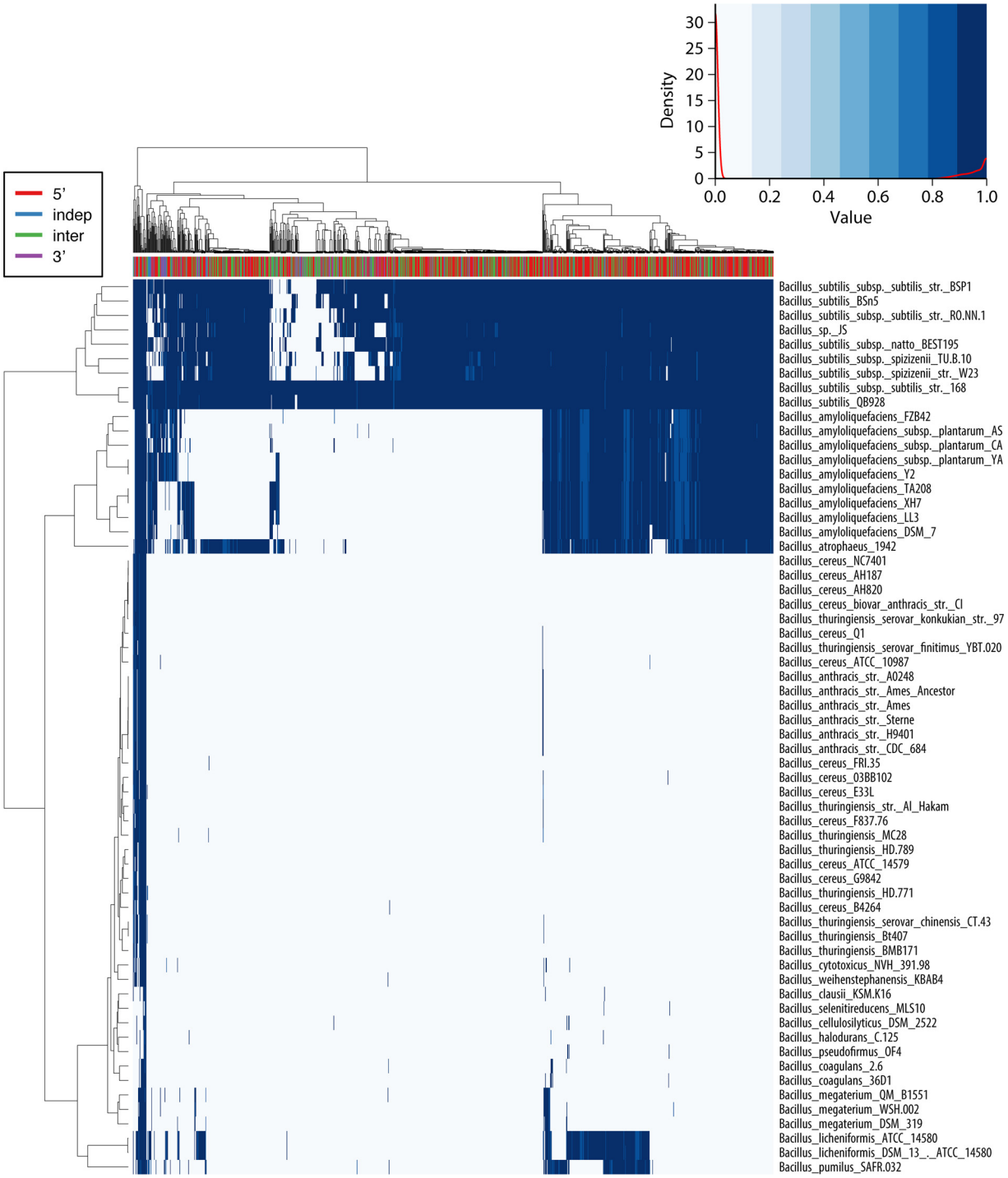

FIG 3 Heat map of sense RNA segment conservation. The proportion of sequence conservation in the indicated genomes was plotted for each sense RNA segment. Rows and columns were reordered automatically using bidirectional hierarchical clustering. The color scale on top indicates the category of each investigated RNA segment. The evolutionary conservation of all RNA segments reported by Nicolas et al. (15) was analyzed using the blastall program (BLASTN 2.2.26). Sequence comparisons were performed against the 62 *Bacillus* genomes available in GenBank (as of 31 January 2013) with the default blast v2.2.26 parameters, except for the filtering of low-complexity regions, which was disabled. The results were tabulated with the fraction of identity and the coordinates of the homologous regions as output (see Table S3 in the supplemental material).

extent of translation initiation, prevent RNA degradation, or control transcription through transcriptional attenuation (TA).

RNA structures in the 5′ region of an mRNA are known to influence the extent of translation, especially when they are located in the close vicinity of the ribosome-binding site (RBS) (48–50). Particular examples of how 5′ RNA structures can impact translation, for example via sequestration of an RBS, are discussed below, especially in the section on riboswitches.

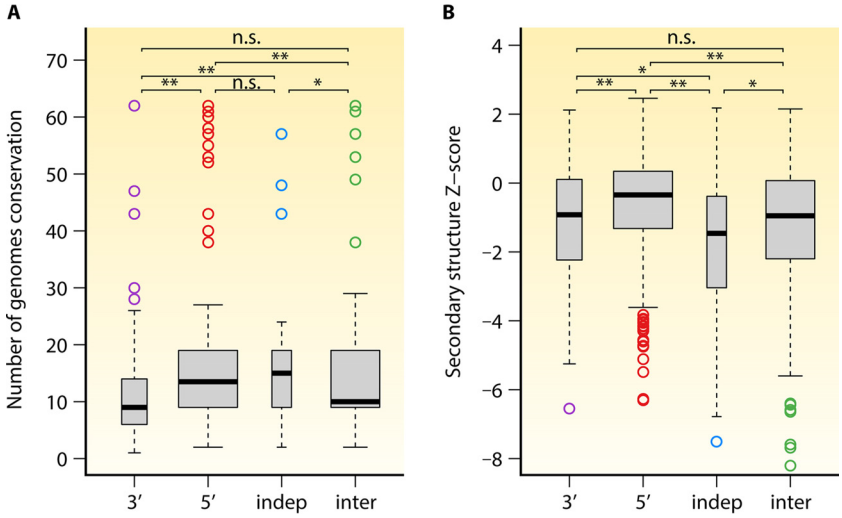

FIG 4 Independent RNA segments are the most conserved and structured RNA segments of *B. subtilis*. Box plots of species-level conservation (A) and predicted secondary structure MFE Z-score (B) of sense RNA segments per category. Significance was tested using analysis of variance with the Tukey honestly significant difference test at 99% confidence. *, significance with *P* value of ≤0.05; **, significance with *P* value of ≤0.01; n.s., not significant. The secondary structure of all RNA segments reported by Nicolas et al. (15) was predicted using the source code of the RNAfold program from the Vienna RNA package (46). The minimum free energy (MFE) output values were transformed into MFE Z-scores (number of standard deviations from the mean). For each segment, the mean and the standard deviation of the score distribution were computed based on 100 shuffled versions of the original sequence, thereby accounting for the length and mononucleotide composition of the segment.

RNA structures can also have a regulatory effect on RNA degradation by RNases. In brief, two different classes of RNases can be distinguished: those that cleave within an RNA sequence (endoribonucleases) and those that remove nucleotides, one at a time, from either the 5′ or 3′ end (exoribonucleases). *B. subtilis* nucleases are present in a multiprotein complex called the RNA degradosome (51) in which the global endoribonuclease is RNase Y and the global exoribonuclease is RNase J1 (52). Since RNase J1 progresses in the 5′ to the 3′ direction after removal of the 5′ pyrophosphate by RppH, structural elements in the 5′ region can block RNase progression along the transcript (49, 52–54). Another endoribonuclease, RNase III, specifically cleaves double-stranded stretches of RNA (53). For a recent review of RNA turnover in Gram-positive bacteria, please refer to the work of Durand et al. (55). Multiple *B. subtilis* 5′UTRs have been reported to have an effect on mRNA stability due to strong stem-loop structures directly at the 5′ end of the mRNA (49). Conceivably, these structures could affect the level of ribonucleolytic attack. These 5′UTRs can be small, with some of the examples discussed by Sharp and Bechhofer having a size of around 30 bases (49). However, 5′UTRs below a size of 50 nucleotides (nt) were not annotated as separate RNA segments in the study by Nicolas et al. (15). One of the examples discussed by Sharp and Bechhofer is the 5′UTR of *aprE* (49), which is larger than 50 nt and was annotated as S363 (MFE Z-score −1.27). The observation that 5′UTRs do not have to be very long to have a strong effect on mRNA stability raises the

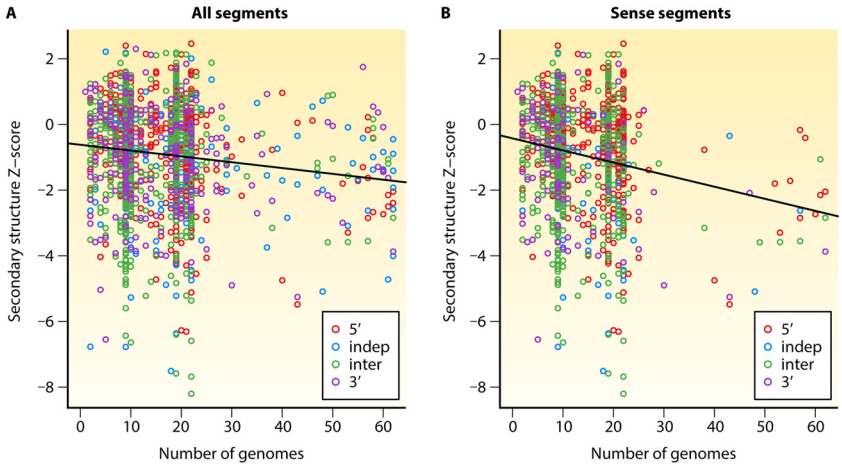

FIG 5 Relationship between conservation and predicted secondary structure. (A) Plotting of the MFE Z-scores for predicted secondary structures as a function of feature conservation in different genomes indicates a trend toward more predicted secondary structure (i.e., a lower MFE Z-score) in the more conserved segments. (B) The trend toward more predicted secondary structure (i.e., a lower MFE Z-score) becomes stronger when only sense RNAs are considered.

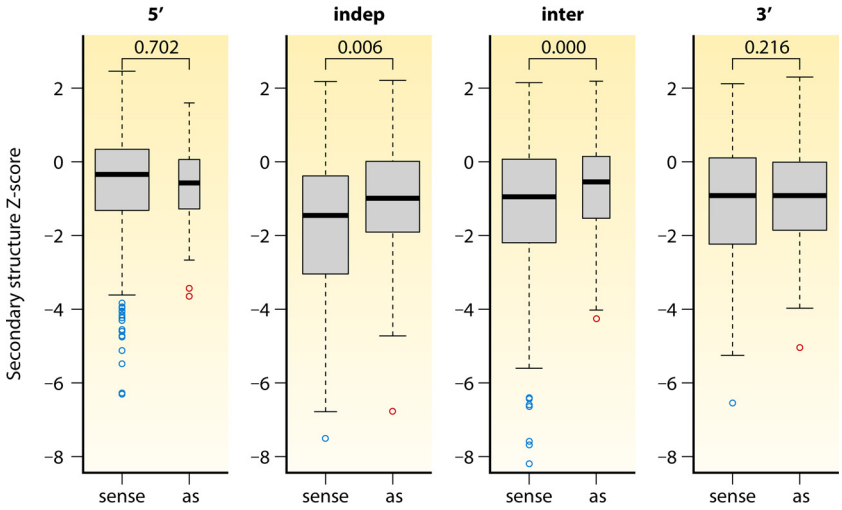

FIG 6 Sense-independent and intergenic RNA segments are more structured than their antisense counterparts. Comparisons of the predicted secondary structure MFE Z-scores of the sense and antisense RNA segments belonging to the four RNA segment categories are shown in box plots. Statistical comparisons were performed with a Welch two-sample *t* test, and the resulting *P* values are indicated. The sense-all-independent and sense-all-intergenic segments are significantly more structured than their antisense counterparts.

question of why many 5′UTRs identified by Nicolas et al. are so much longer. Do these 5′UTRs have a function in mRNA stability regulation as well? Or are they perhaps platforms for protein or other RNA factors to bind? In this context, it should be noted that some of these 5′UTRs represent previously reported leader regions that had not yet been annotated in the *B. subtilis* 168 genome annotation used for the study by Nicolas et al. (see Table S1 in the supplemental material).

Transcriptional attenuation (TA) systems are 5′-*cis* regulatory elements that can fold into two alternative structures, one of them representing an intrinsic terminator (56). Intriguingly, a multitude of variations in TA systems is nowadays known in which attenuation is modulated by, for instance, the binding of metabolites, tRNAs, or specific regulatory proteins or by the translation of a so-called leader peptide that influences the timing and outcome of the RNA folding process. A genome-wide approach to identify TA sites was recently described, which is based on the accurate mapping of the 3′ ends of RNA molecules through RNA-seq (57). In this approach, named termSeq, RNA was isolated from bacteria that were either exposed to antibiotics or not, and the isolated RNAs were then processed to specifically map their 3′ ends. Subsequently, those sequences showing read-through transcription in the presence of antibiotics were selected. Using this approach, 82 possible TA sites were identified in *B. subtilis* 168, including 18 novel sites. Interestingly, all of these TA sites map to S-segments included in our present analysis (see Table S1 in the supplemental material), and for some, the termination signals are clearly visible in the tiling array analyses described by Nicolas et al. (15, 35). For instance, this applies to *bmrB*/*yheJ*, S1228, S1031, and S1276 as referred to below and documented in the expression browser at http://genome.jouy.inra.fr/seb.

The so-called ribosome-mediated TA is possible in prokaryotes since transcription and translation are coupled. A recently documented example of this type of TA in *B. subtilis* concerns the regulation of the *bmrCD* operon, encoding a multidrug-associated ABC transporter (50, 57). In this study, it was shown that unimpaired translation of the leader peptide BmrB, which is encoded within the upstream structured RNA, prevents transcription of *bmrCD* under normal conditions. However, when translation of BmrB is slowed down by the addition of ribosome-targeted antibiotics, an antiterminator structure forms that allows transcription to continue into the *bmrCD* operon (50). Interestingly, another TA system was reported in the 5′UTR of *vmlR*, which is also responsible for the efflux of ribosome-targeting antibiotics in *B. subtilis* (58). The 5′UTR of *vmlR* S198 also contains multiple putative ORFs, but it remains to be investigated whether these ORFs are involved in TA and to what extent the respective TA regulation mechanism is similar to that of *bmrCD* (50).

Out of the 1,583 new RNA segments identified by Nicolas et al., the 5′UTRs are the most dominant category (676 segments, 43% of the total). This was also the case in the RNA-seq study by Irnov et al. (14), where 40 long 5′UTR leader regions were identified of which 38 are also part of the set of 676 5′UTRs identified by Nicolas et al. (15). In addition, the large extent of transcriptional regulation via 5′UTRs in *B. subtilis* is reflected by the 114 previously identified or proposed regulatory RNAs in 5′UTRs, which represent a far greater number than the total number of regulatory RNAs from all other groups combined (see Table S1 in the supplemental material). Overview plots of the GC%, conservation, secondary structure, and length of the new RNA segments annotated as 5′UTRs reveal a large diversity in this 5′UTR group (Fig. 7). Only a relatively small percentage (13%) of these 5′UTRs are antisense segments. What distinguishes these 5′ asRNAs from the other segments is their length, since they are much longer than the 5′ sense RNAs. The three most conserved sense 5′UTRs are S17 (62/62 genomes), S55 (61/62 genomes), and S49 (60/62 genomes). S17 precedes the BSU_misc_RNA_2 upstream of the *scr* gene for 4.5S RNA, while S55 and S49 correspond to the S10 and S11 leaders of the ribosomal protein genes *rpsJ* and *rplK*, respectively.

In the following paragraphs, three classes of well-studied RNA switches situated in 5′UTRs will be discussed. The first class is composed of RNA switches that respond to the binding of metabolites, the second contains RNA switches that control the transcription of amino acid-related genes in response to the cellular

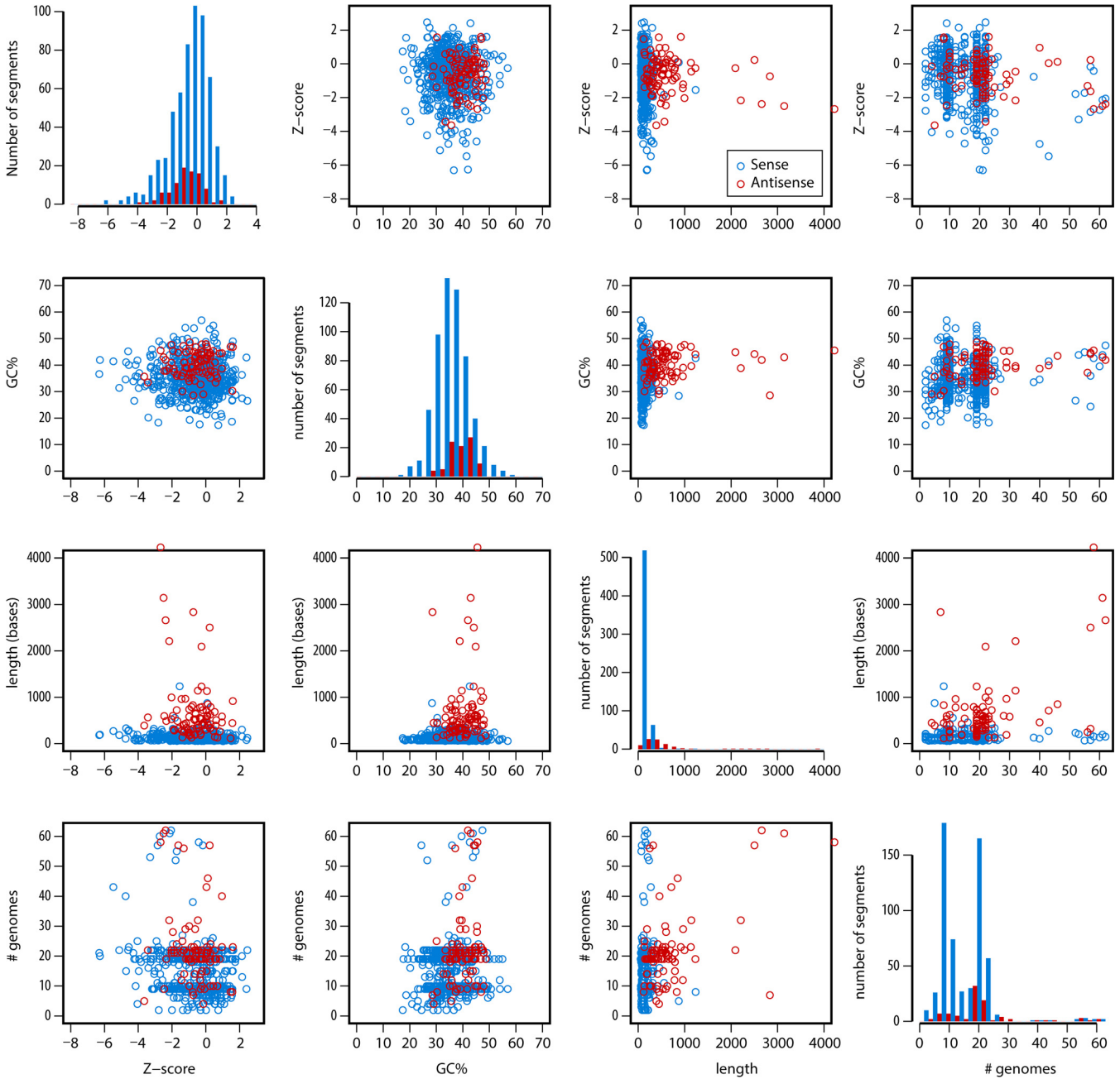

FIG 7 Overview plot of the predicted secondary structure MFE Z-scores, GC%, lengths, and species-level conservation for RNA segments of the 5′UTR category. MFE Z-scores were used as a measure for the stability of the predicted secondary structure of the S-segments. MFE Z-scores were computed from the RNAfold minimum free energy (MFE) of the S-segment, minus the mean MFE of 100 shuffled sequences with the same length and nucleotide composition, divided by the standard deviation of the MFE between the 100 shuffled sequences. GC%, percent guanine + cytosine bases in the respective S-segment sequence; length, S-segment sequence length; # genomes, number of genomes in which a significant blastall (BLASTN 2.2.26) hit for the segment was identified (default settings but filtering of low-complexity regions disabled). The 62 *Bacillus* genomes available in GenBank (as of 31 January 2013) were used.

ratio of charged and uncharged tRNA molecules, and the third includes protein-binding RNA switches. Last, 5′UTRs with other or unknown functions will be discussed.

Metabolite-Binding Riboswitches

Metabolite-binding riboswitches are fascinating RNA regulatory elements that can directly integrate metabolite concentrations in transcriptional regulation by controlling levels of their downstream gene(s). These downstream genes are often functionally related to the metabolism of the respective bound metabolite. Riboswitches generally consist of two parts, a ligand-binding aptamer part, which changes conformation upon ligand binding, and a signal transduction part, called the expression platform (29). This expression platform regulates gene expression by adopting alternative RNA structures that affect transcription or regulation. These structural changes often entail the formation of alternative hairpin structures, which can, for instance, create or disrupt

transcriptional terminators or antiterminators. In addition, these structural rearrangements can occlude or expose ribosome-binding or RNase cleavage sites. Because of their modularity and catalytic diversity, riboswitches are interesting from an evolutionary perspective as ancient molecules from the RNA world (29), and they could be employed as metabolite-sensing parts in synthetic biology designs (59, 60).

B. subtilis contains approximately 35 riboswitches that, in total, regulate ~2% of all genes present in this bacterium (2, 61, 62) (see Table S1 in the supplemental material). Among other ligands, there are for instance riboswitches for flavin mononucleotide (FMN), lysine, guanine, *S*-adenosylmethionine (SAM), thiamine pyrophosphate, and the second messenger molecule cyclic-di-AMP (29, 57, 61, 63, 64) (see Table S1 in the supplemental material). Riboswitches display a wide diversity in their regulatory mechanisms, and even riboswitches of the same class can regulate their downstream genes in alternative ways. This is illustrated by the example of the FMN-box riboswitch, which responds to binding of the flavin mononucleotide. Specifically, the FMN-box inhibits transcription of the *ribD* gene via the formation of a terminator structure, whereas translation of *ribU* is inhibited via sequestration of the RBS (65).

Since the number of identified *cis*-acting riboswitches was found to be higher in Gram-positive bacteria than in Gram-negative bacteria, and the contrary seemed to be the case for sRNAs, this has led to the suggestion that Gram-positive bacteria rely more on transcriptional than posttranscriptional regulation (2, 14). However, since recent transcriptomics studies (14, 15) have identified many independently expressed putative regulatory RNAs in *B. subtilis*, this suggestion may have been premature, and it may thus have simply reflected a bias in the direction of study. In this context, it is noteworthy that Nicolas et al. did not identify any new metabolite-binding riboswitches in *B. subtilis* (15). This probably reflects the previously performed extensive genome-wide screens for riboswitches and the relative ease of riboswitch discovery, which is facilitated by their structural conservation (62, 66). Notably, six metabolite-binding riboswitches that were reported previously were not annotated as such in the genome sequence that was employed by Nicolas et al. (15). This led to the annotation of these known riboswitches as "new" S-segments (see Table S1 in the supplemental material).

tRNA-Controlled RNA Switches

The tRNA-controlled RNA switches regulate the transcription of genes encoding amino acid metabolism and genes controlling the charging of their respective tRNA by means of the T-box mechanism (31, 67). Each T-box RNA has evolved to respond to a specific uncharged tRNA. Most T-box RNAs can fold into two alternative secondary structures. The energetically most favorable structure prevents transcription or translation, while the other one allows transcription and translation of the downstream genes. Preventing transcription is achieved by an intrinsic terminator, and preventing translation is performed by an RBS-binding helix. The formation of both these elements is energetically more favorable than the formation of the transcriptional antiterminator. The antiterminator of the T-box RNA is stabilized only by binding of its cognate uncharged tRNA. This discrimination between uncharged and charged tRNAs is mediated by the specific pairing of the four unpaired nucleotides at the 3′ end of the tRNA with the antiterminator. In charged tRNAs, these nucleotides are prevented from binding by the presence of the amino acid. Details of these elegant regulatory systems are comprehensively reviewed by Gutíerrez-Preciado and colleagues (31). Altogether, we identified 19 T-box RNAs in *B. subtilis* 168 (see Table S1 in the supplemental material). The majority of these, 14 in total, were already annotated in the genome sequence used by Nicolas et al. (15), and for these RNAs, we followed the "BSU_misc_RNA" nomenclature. For the five others, we have now suggested an S-segment nomenclature as presented in Table S1 in the supplemental material.

Protein-Binding RNA Switches

Protein-binding RNA switches are RNA structural elements that control the expression of downstream genes based on the binding of a specific regulatory protein. The mechanisms used in these switches are again rather diverse, but generally, they involve either the attenuation of transcription or the inhibition of translation initiation. We listed 18 different RNA-binding proteins of 7 different types, for which the respective RNA switches are situated upstream of 28 different genes or operons (see Table S1 in the supplemental material). Of these 28 genes, 6 are preceded by so-called BSU_misc_RNAs, which were already defined in the previous genome annotation; 14 have been attributed an S-number by Nicolas et al. (15); and 8 are not linked to any of the new S-segments (see Table S1 in the supplemental material).

Eight of the 28 protein-binding RNA switches are responsible for autoregulation of ribosomal proteins L10, L13, L19, L20, L21, S4, S10, and S15. The most obvious function of these regulatory elements is to prevent the presence of large numbers of unbound ribosomal proteins by inhibiting transcription or translation of the downstream genes (32, 68) (see Table S1 in the supplemental material). Two of these switches are present in front of genes for the so-called primary-binding proteins S4 and S15 of the small ribosomal subunit and could thereby directly control ribosome levels (69, 70).

The second group of protein-binding RNA switches belongs to the so-called BglG family. These RNA switches are situated in the 5′UTRs of genes responsible for alternative sugar utilization (30) (see Table S1 in the supplemental material), and there are four different RNA-binding proteins involved in their regulation. These are LicT with three known targets, SacY with two targets, and GlcT and SacT with one target each (see Table S1 in the supplemental material). These proteins are differentially phosphorylated based on activity of the phosphoenolpyruvate:sugar phosphotransferase system (PTS). Because of the relatedness of these switches there is the potential for cross talk between the different regulatory elements. However, it was found that specific differences have evolved that allow the regulatory systems to distinguish between the involved binding proteins (30). A well-known example of regulation by the BglG family of RNA switches is the LicT-mediated regulation of *bglP* transcription. In the absence of glucose, LicT is present in its active phosphorylated state, where it mediates antitermination via binding to structural elements in S1513, thereby facilitating read-through into *bglP*. In the presence of glucose, LicT is not phosphorylated by HPr and cannot bind S1513. This allows for the terminator structure to form, which prevents read-through into *bglP* (30).

A third group of protein-binding RNA switches is formed by those binding the *trp* RNA-binding attenuation protein (TRAP), encoded by the *mtrB* gene. Active TRAP regulates tryptophan metabolism by a variety of TA and translation control mechanisms

(71). TRAP assembles into a ring-shaped oligomer of 11 subunits that is allosterically activated by tryptophan to bind specific RNA sequences in the presence of tryptophan (71). Interestingly, an anti-TRAP protein that inhibits TRAP is encoded by *rtpA*, and this gene is preceded by a tryptophan-responsive T-box RNA regulatory system (71) (see Table S1 in the supplemental material). This combination of allosteric activation of TRAP and the complete silencing of TRAP activity using anti-TRAP allows for fine-tuning of the biosynthesis of the costly amino acid tryptophan.

RNA chaperones that are activated upon cold stress are involved in the fourth group of protein-binding RNA switches. This regulatory mechanism is exemplified by the CspB and CspC proteins of *B. subtilis*, which are important for the survival of cold shock and autoregulate their own transcription by acting as antitermination proteins (72, 73). While the molecular details of the actual regulatory mechanism remain to be elucidated, it seems that the upstream secondary structures (S179 and S315) are functionally analogous to RNA thermometers that allow antitermination only at low temperatures.

The final three proteins known to bind RNA switches in *B. subtilis* are PyrR, GlpP, and Hut (see Table S1 in the supplemental material). In the absence of uridine nucleotides, PyrR antiterminates transcription of genes responsible for biosynthesis and uptake of these nucleotides (74, 75). Similarly, GlpP antiterminates transcription in response to glycerol-3-phosphate levels (76) at three sites in the *B. subtilis* 168 genome (now annotated as S82, S321, and S322; see Table S1 in the supplemental material). HutP regulates the *hut* operon, which is responsible for histidine utilization, by an antitermination mechanism sensitive to histidine levels (77).

5′ Untranslated Regions with Other or Unknown Functions

Four other 5′UTRs that are not RNA switches have been characterized in *B. subtilis*. The first example of such 5′UTRs is the aforementioned 5′UTR of *aprE*, which is annotated as S363. The second is a long transcribed sequence triplication preceding the *asnH* operon, which was shown to be responsible for mRNA stabilization (78). Interestingly, in the study by Nicolas et al. this 5′UTR is annotated as an independent segment, S1534, because under some conditions S1534 is highly expressed while the downstream genes are not. The conditions with the largest expression difference between S1534 and its downstream gene *yxbB* are stationary-phase growth in lysogeny broth (LB), a late stage in sporulation (S6), salt stress, and confluent growth on plates (BC). Transcriptional read-through beyond S1534 takes place under most other conditions. It could thus be that this phenomenon is caused by a currently unreported condition-specific control of read-through transcription mediated by S1534. A third example is the 5′UTR of *abnA*, which was named S1087. The *abnA* mRNA has a long half-life that is possibly conferred by its UTR (49, 79), but the underlying mechanism has not yet been reported. The fourth example concerns the 5′ leader region of *cwlO*, S1327 (80). CwlO is a D,L-endopeptidase-type autolysin involved in the highly regulated process of cell wall turnover. Noone et al. reported that the *cwlO* transcript is highly unstable due to RNase Y-dependent cleavage in this leader region (80). It was proposed that this instability of the *cwlO* mRNA is of physiological relevance as it would give the transcriptional regulation via the WalRK two-component signal transduction system a more direct impact on the CwlO protein levels (80).

The following paragraphs will first detail the types of structural elements in the 5′ regions of mRNAs that are known to have an effect on mRNA stability, and subsequently, the 10 most structured and conserved 5′UTRs identified among the new 5′UTR segments will be discussed.

Sharp and Bechhofer (49) distinguished five features of the 5′ region that may affect mRNA stability: (i) strength of the RBS sequence, (ii) distance from the 5′ terminus to the RBS sequence, (iii) distance of the RBS from the 5′-terminal secondary structure, (iv) unpaired 5′ nucleotides, and (v) the strength of the 5′-terminal secondary structure. First, elongating ribosomes along an mRNA protect the sequence from exposure to RNases (81). A strong RBS is thus expected to lead to better protection of the mRNA than a weak one. However, RBS strength by itself is only a determinant for mRNA stabilization when ribosome binding occurs in close proximity to the 5′ terminus (49). This was concluded from the finding that only mRNAs with RBS sequences close to the 5′ end are stabilized by increasing RBS strength (49), and it is most likely due to the fact that mRNA degradation in *B. subtilis* often proceeds in the 5′-to-3′ direction or is initiated at the 5′ end. This is probably also the reason why the distance from the 5′ terminus to the RBS sequence is a second feature of the 5′ region that can affect mRNA stability. Notably, this distance was of no influence on mRNA stability when the secondary structure at the 5′ end was relatively strong, which relates to the third feature determining mRNA stability. Nonetheless, when the initiation codon for translation was mutated in an mRNA with a strong 5′ secondary structure, the mRNA half-life was still reduced (49). The latter finding is consistent with the results from studies on the 5′UTR S363 preceding *aprE* and an artificial stem-loop preceding *rpsO* (82), where the stem-loop and ribosome binding were found to be more important than translation (83).

With regard to the distance of the RBS to the 5′-terminal secondary structure, it seems that sufficient spacing between these two is required for mRNA stabilization since strongly binding ribosomes can disrupt the secondary structure in the 5′-terminal sequence. However, the reverse can also be true, where secondary structure elements prevent ribosome binding and thereby inhibit translation (84). Beyond the influence of ribosomes on mRNA stability, the presence of seven unpaired nucleotides at the 5′ end of an mRNA was shown to form an initiation site for degradation by RNase, thus representing a fourth feature that determines mRNA stability (49). Besides this, RNase progression can probably still be inhibited by a strong 5′ structure, which corresponds to the fifth feature determining mRNA stability. A series of mutations in a model 5′-terminal structure revealed that this is potentially the most important determinant for mRNA half-life (49). In this particular study, a minimum structure was identified that was sufficient to confer mRNA stability, and further increasing the MFE of this structure had no additional stabilizing effect (49).

To pinpoint the potentially most relevant 5′UTRs of *B. subtilis*, the 20% most structured (lowest MFE Z-score) and 20% most conserved RNA segments from the 5′UTR category were selected and the segments that were present in both lists were sorted based on their MFE Z-scores. The first 10 segments of this list (Table 1) were manually analyzed for the presence of known regulatory elements using RibEx, an online tool for identifying riboswitches and other conserved bacterial regulatory elements (38). The most structured and conserved 5′UTR thus identified is S1501 (21/62 genomes, MFE Z-score −6.3), which precedes *yxjB*, a gene of un-

TABLE 1 Top-10 selection of RNA segments from each category[a]

Name	No. of species	Classification	Sigma factor	Z-score	MFE centroid	Length (nt)	GC%
S1501	21	5′	SigA	−6.3	−65	186	37
S1513	20	5′	SigA	−6.3	−57	193	42
S304	43	5′	SigA	−5.5	−79	273	41
S1441	22	5′	SigA	−5.1	−47	188	31
S1228	40	5′	SigA	−4.8	−26	110	35
S1563	15	5′	SigA	−4.7	−98	331	41
S1276	22	5′	SigA	−4.6	−95	270	45
S1181	23	5′	SigA	−4.6	−67	292	40
S1031	21	5′	SigA	−4.3	−44	207	34
S476	19	5′	SigH	−4.2	−21	83	27
S347	43	3′	SigK	−5.3	−72	252	38
S175	30	3′	SigA	−4.9	−72	317	37
S127	22	3′PT	NA	−4.2	−46	152	39
S987	17	3′	SigA	−4.0	−96	415	35
S1450	16	3′MT	SigA	−4.0	−610	2,483	44
S1188	62	3′	SigK	−3.9	−78	287	42
S187	11	3′	SigA	−3.9	−103	408	38
S611	25	3′	SigA	−3.8	−121	495	40
S151	10	3′	Sig−	−3.8	−73	244	45
S810	13	3′	SigA	−3.7	−58	265	35
efeN 3′/S1476	22	Inter	SigA	−8.2	−100	282	39
S239	22	Intra	SigA	−7.7	−106	255	43
S1503	19	Inter	SigA	−7.6	−38	104	34
S565	10	Inter	SigA	−6.6	−32	89	31
S600	22	Intra	SigA	−6.6	−29	91	40
S918	19	Inter	SigB	−6.4	−53	163	34
S160	22	Inter	SigWXY	−5.4	−100	328	46
S424	13	Inter	SigK	−5.2	−43	132	38
S1554	21	Intra	SigA	−5.2	−29	81	40
S1524	19	Intra	SigA	−4.8	−43	85	58
as-BsrH/S977	18	Indep	SigA	−7.5	−143	508	35
S357	19	Indep	SigGF	−6.4	−67	297	37
S326	10	Indep	SigGF	−5.3	−31	111	31
RnaC/S1022	19	Indep	SigD	−5.2	−45	126	42
vmlR leader/S198	48	Indep	SigA	−5.1	−58	198	33
bsrE/S717	17	Indep	SigA	−4.9	−55	154	44
FsrA/S512	22	Indep	SigA	−4.8	−24	102	41
CsfG/S547	61	Indep	SigGF	−4.7	−42	124	48
S2	22	IndepMT	SigK	−4.4	−68	274	37
S1292	19	IndepMT	SigA	−4.3	−75	520	36

[a] The 20% most conserved and the 20% most structured RNA segments of each category were selected and sorted based on their MFE Z-scores. The first 10 of each resulting list are presented. The "Classification" column includes the following classes of segments: 5′, 5′UTR; 3′, 3′UTR; 3′PT, 3′UTR with partial termination; 3′MT, 3′UTR without termination; Inter, intergenic regions between genes under the control of distinct promoters; Intra, intergenic regions between genes under the control of one promoter; Indep, independent segment with own promoter and terminator signals; IndepMT, independent segment without clear terminator. NA, not applicable.

known function. On the opposite strand (but not overlapping), there is a known purine riboswitch in front of *yxjA*, and this might point to a function of this genomic region. The predicted large stem-loop structure and the relatively low correlation between S1501 and *yxjB* transcript levels suggest a regulatory role in controlling read-through transcription, and this was confirmed by termSeq (57). The second segment in the top-10 list, S1513 (20/62 genomes, MFE Z-score −6.3), was already reported above in the section on protein-binding RNA switches since it is the LicT-dependent antiterminator in front of *bglP*, which encodes a phosphotransferase system component (85) (see Table S1 in the supplemental material). The third and fourth identified 5′UTRs in the top-10 list, S304 (43/62 genomes, MFE Z-score −5.5) and S1181 (23/62 genomes, MFE Z-score −4.6), precede the 16S RNA genes *rrnD* and *rrnB*, respectively. These locations may actually explain the conservation of S304 and S1181, since the 16S rRNA genes are naturally highly conserved. It is currently unclear whether these elements are responsible for the regulation of 16S rRNA transcription or perhaps for the maturation of the rRNA. The fifth segment, S1441 (22/62 genomes, MFE Z-score −5.1), precedes the pyrimidine biosynthesis gene *pyrG*. This element is known to confer transcriptional termination in the presence of cytidine nucleotides. Specifically, low levels of CTP induce the reiterative addition of guanine residues at the 5′ end of the *pyrG* mRNA. This poly(G) sequence binds with the C- and U-rich downstream terminator, thus preventing attenuation (86, 87) (see Table S1 in the supplemental material). The sixth element is the 110-nucleotide segment S1228 (40/62 genomes, MFE Z-score −4.8), which precedes the *dhbA* gene, containing a biosynthesis gene for the siderophore bacillibactin (14). The transcription of

S1228 is highly correlated with that of *dhbA* (0.98) (15). Interestingly, the *dhbA* transcript levels were previously reported to be decreased upon RNase Y deletion (88). Thus, it is possible that the three strong stem-loops in the predicted S1228 structure are responsible for controlling the *dhbA* mRNA levels, and it would be interesting to study the role of RNase Y in this process. Likewise, the *yhrE* gene, encoding formate dehydrogenase, is also stabilized upon RNase Y depletion (88), and its 5′UTR S1031 (21/62 genomes, MFE Z-score −4.3) also contains three strong stem-loops. S1031 is in fact the seventh segment in the top-10 list of predicted 5′UTRs and was also identified by termSeq (57). The eighth 5′UTR in this list is S1563 (15/62 genomes, MFE Z-score −4.7), the expression of which is highly correlated (0.96) with its downstream gene *yybP*. Although the function of *yybP* is unknown, its overall expression profile suggests a role in sporulation. Further, *yybP* is induced in the presence of Mn^{2+}. The 331-nucleotide segment S1563 contains a previously reported conserved putative riboswitch element with unknown function referred to as *yybP/ykoY* (62) (see Table S1 in the supplemental material). Interestingly however, S1563 also contains an ORF that can encode a protein of 101 amino acids. It remains to be seen whether this hypothetical protein is involved in the manganese-mediated regulatory mechanism and/or could have an independent function. The ninth considered 5′UTR is S1276 (22/62 genomes, MFE Z-score −4.6). This 5′UTR precedes *yvrC*, and it contains a previously reported cobalamin riboswitch (89), which is generally associated with vitamin B_{12} metabolism genes, which might point to a function for *yvrC* in vitamin B_{12} metabolism. Nevertheless, *yvrC* is currently annotated as being similar to an iron-binding protein component of an ABC transporter. S476 (19/62 genomes, MFE Z-score −4.2) is the 10th considered 5′UTR. This highly structured 5′UTR of 83 nucleotides precedes the highly expressed *ykoM* gene. YkoM is a putative transcriptional regulator from the MarR family. Conceivably, S476 has a role in mediating *ykoM* mRNA stability and abundance, but this has not yet been investigated.

3′ Untranslated Regions

3′UTRs are RNAs that are cotranscribed downstream of a protein-encoding gene. Because of this location, 3′UTRs can originate from incomplete termination after transcription of a CDS. Their regulatory functions are widely acknowledged in eukaryotes, but it is generally believed that direct regulation of gene expression via structured elements in the 3′UTR is highly unusual in bacteria (66). This is not to say that the 3′ end is not important for regulatory RNA molecules, since structured parts are believed to confer stability and provide a platform for interactions. Of note, miRNAs in eukaryotes generally regulate mRNA levels via the 3′UTR, and it has been found that this represents a selective pressure on 3′UTRs (90). Such 3′-end regulation contrasts with the 5′-end regulatory model proposed for prokaryotes. This difference in regulation between pro- and eukaryotes is conceivably due to differences in the mechanisms of RNA degradation in these two domains of life. The processing and degradation of mRNA seem more complex in eukaryotes and often take place from the 3′ end by the exosome after removal of the poly(A) tail (91, 92), In contrast, RNA decay in *B. subtilis* is initiated endonucleolytically by RNase Y or at the 5′ end via dephosporylation and subsequent 5′-to-3′ exoribonucleolytic degradation by RNase J1 (52–54, 93). This suggests that signal integration via posttranscriptional regulation at the 3′UTR is probably not as effective in prokaryotes as it is in eukaryotes. Interestingly, RNA segments that appear to be 3′UTRs could (additionally) function as *trans*-acting small regulatory RNAs. This may be the case when the RNA is processed, or when transcription initiation takes place within the protein-encoding sequence and the mRNA and sRNA share a terminator (16). Out of all S-segments (15), 249 (16%) are 3′UTRs; 78 of these arise from partial transcriptional termination and 46 arise from no termination. Thus, 3′UTRs in *B. subtilis* are relatively rare compared to 5′UTRs, and they often arise from incomplete control of transcriptional termination. The latter aspect has also been linked to the high percentage (57%) of asRNAs found in the 3′UTR category.

The proposed lower importance of 3′UTRs in bacterial post-transcriptional regulation is supported by their general statistics. The sense 3′UTRs are much less conserved than the sense 5′UTRs and independent segments (Fig. 4A). Overview plots of the GC%, conservation, secondary structure, and length of the new RNA segments annotated as 3′UTRs reveal that 3′ asRNAs are much longer on average. Unsurprisingly, these 3′ asRNAs tend to be more conserved, because they share sequence conservation with their cognate protein-encoding sequences (Fig. 8 and 9). The three most conserved sense 3′UTRs are S1188 (62/62 genomes) downstream of the sporulation-induced putative cysteine dioxygenase *cdoA/yubC*, S18 (47/62 genomes) downstream of the *scr* gene for 4.5S RNA, and S347 (43/62 genomes) downstream of the small spore protein-encoding gene *sscA/yhzE*.

To date, nine 3′ regulatory elements have been described in *B. subtilis*: two by Alen and Sonenshein (94) and seven by Rasmussen et al. (35). The first two were shown to be specific binding platforms for aconitase, which is encoded by the *citB* gene (94). Interestingly, these two sequences are similar to iron-response elements of eukaryotic origin, located at the 3′ end of the *qoxD* gene encoding the major cytochrome oxidase and between the *feuA* and *feuB* genes that encode components of an ABC transporter for iron uptake. Because of their small size, these two elements were not annotated as 3′UTRs by Nicolas et al., and if they had been annotated, the one between *feuA* and *feuB* would have been classified as an intergenic segment. The seven 3′ segments described by Rasmussen et al. (35) are paralogues at the level of sequence and structure, since they are all predicted to fold into a similar strong secondary structure. In fact, the MFE Z-score S1476, the 3′UTR of *efeB/efeN/ywbN* (95), is the lowest of all segments, −8.2. It is presently unclear what the function of this RNA structure is, but it has been speculated that it is linked to the localization of the cognate mRNA transcripts since most of these genes encode membrane-related proteins (35). Of these seven RNA segments, four are part of the complete set of RNA segments considered here. The absence of the other three is due to their homology, since highly related sequences that could not unambiguously be mapped on the genome were masked for the analysis (15). Of the four considered segments, two are annotated as intergenic segments (S239 downstream of *dagK* and S1476 downstream of *efeB/efeN/ywbN*) since there was no significant transcriptional downshift identified before the start of transcription of the downstream gene. The other two (S1160 downstream of *ytvA* and S127 downstream of *tcyC*) are annotated as 3′, but both were truncated compared to the sequence reported by Rasmussen et al. (35), again due to the exclusion of highly related sequences.

The 20% most structured (lowest MFE Z-score) and 20% most

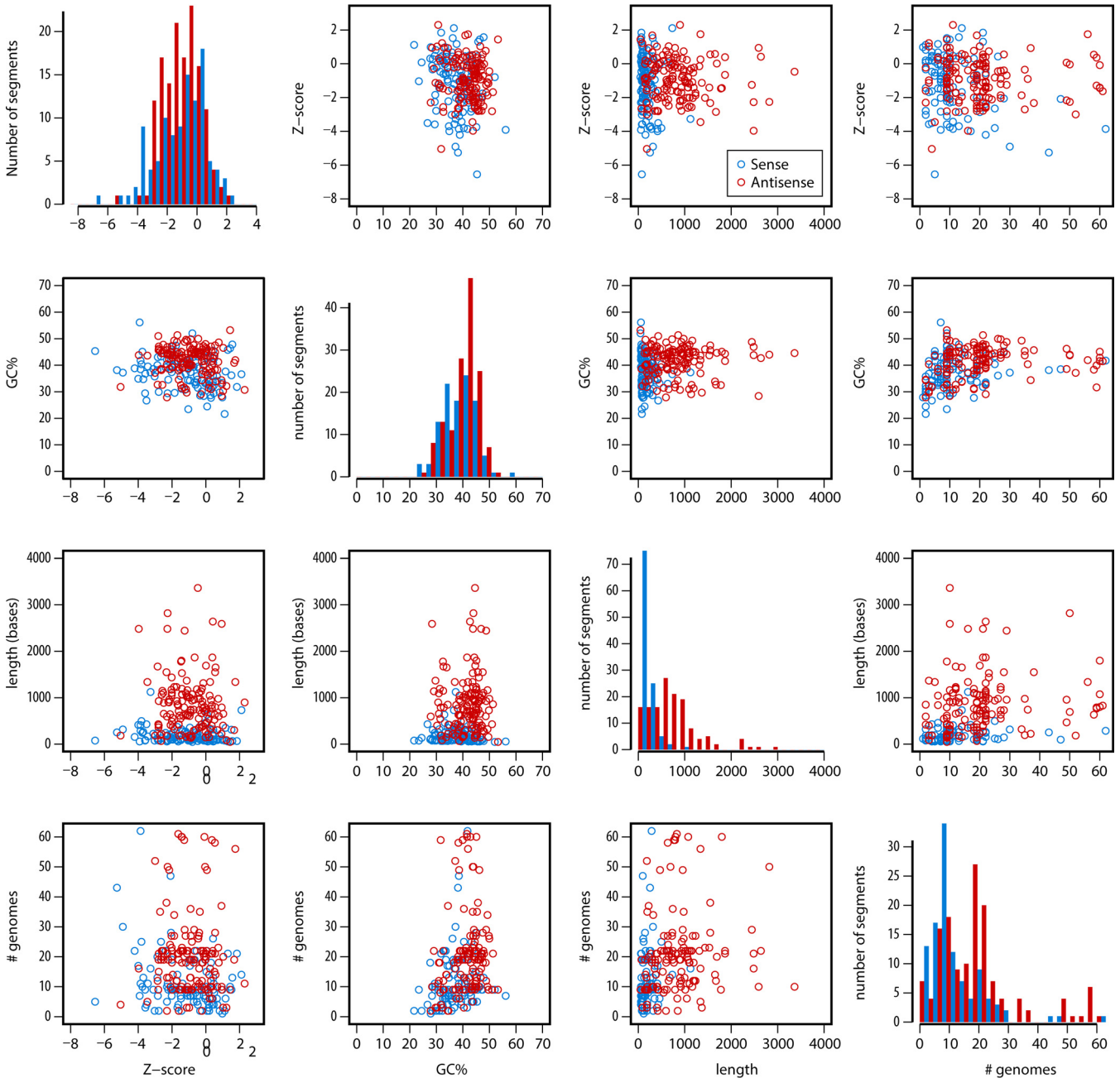

FIG 8 Overview plot of the predicted secondary structure MFE Z-scores, GC%, length, and species-level conservation for RNA segments of the all-3′UTR category. For details on the calculation of MFE Z-scores and abbreviations, see the legend to Fig. 7.

conserved segments from the all-3′ category were selected and sorted by their MFE Z-scores. The first 10 of this list (Table 1) were manually inspected for known regulatory elements using RibEx (38). Judged by their condition-specific expression profiles, two of these 10 most-conserved and structured 3′UTR segments, S175 (30/62 genomes, MFE Z-score −4.9) and S810 (13/62 genomes, MFE Z-score −3.7), may function as independent regulators. This could be a second function in addition to their putative 3′UTR regulatory function, because a transcriptional upshift seems visible just upstream of these segments and because the correlation between their expression and the respective upstream genes *yddN* (0.82) and *yokL* (0.69) is relatively limited. The expression of two other 3′UTRs from the top-10 list, namely, S347 (43/62 genomes, MFE Z-score −5.3) and S1188 (62/62 genomes, MFE Z-score −3.9), is highly correlated with that of their upstream mRNAs, which are almost solely expressed under sporulation-inducing conditions (*yhzE*/*sscA* and *yubC*/*cdoA*). Because of this high expression correlation, their functions are likely tied to the function of their cognate genes. A fifth selected segment, named S151 (10/62 genomes, MFE Z-score −3.8), is most likely an intergenic regulator of the *ydbIJK* operon of unknown function, as judged by the respective expression profiles from the work of Nicolas et al.

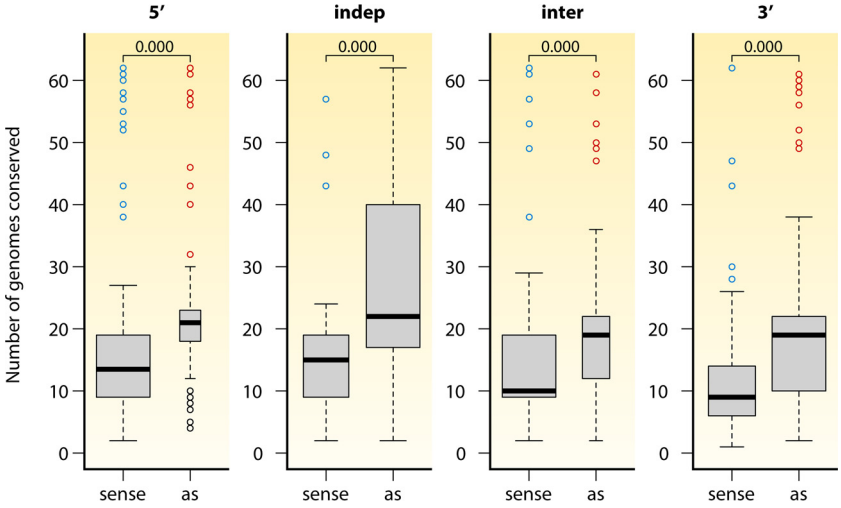

FIG 9 Antisense RNA segments are more conserved at the sequence level than the sense RNAs. Comparisons of the species-level conservation between the sense and antisense RNAs belonging to the four RNA segment categories are shown in box plots. Statistical comparisons were performed with a Welch two-sample *t* test, and the resulting *P* values are indicated. In all cases, the antisense RNAs are more evolutionarily conserved than their sense RNA counterparts. However, this is linked to the sequence conservation of the protein-encoding genes on the sense strand.

(15), and the termSeq study by Dar et al. (57). The sixth 3′UTR segment in the top-10 list, S1450 (16/62 genomes, MFE Z-score −4.0), is annotated as an asRNA that arises from incomplete termination of the *albABCDEFG* operon, which includes the biosynthesis genes for the antilisterial bacteriocin subtilosin. However, the overlapping genes on the opposite strand (*ywhKL*) are hardly ever expressed under the 104 conditions tested by Nicolas et al. (15). In fact, the *ywhKL* genes were expressed only under biofilm-inducing conditions, which argues against an antisense regulatory function of S1450. The seventh top-10 segment is S611 (25/62 genomes, MFE Z-score −3.8), the 3′UTR of the gene of unknown function *ymzD*, which is overlapped by an asRNA in the form of the independent segment S612. However, there is only a very slight transcriptional anticorrelation between these two segments (−0.04), and this argues against antisense regulation of *ymzD* by S612. The last three top-10 segments are S127 (22/62 genomes, MFE Z-score −4.2) downstream of the *tcyABC* operon encoding a cysteine ABC transporter, S987 downstream of the *yrkL* gene encoding an NAD(P)H oxidoreductase homologue (17/62 genomes, MFE Z-score −4.0), and S187 downstream of the *nap* gene encoding the carboxylesterase NP involved in lipid degradation (11/62 genomes, MFE Z-score −3.9).

Intergenic Regions

RNA regions between two protein-encoding genes that are cotranscribed with these genes can control relative mRNA abundance or protein expression of each ORF by regulating posttranscriptional processes, such as mRNA processing, transcriptional termination, or translation initiation (48). It has been shown that bacterial intergenic regions can have strong regulatory effects on their downstream genes, and this has been exploited for fine-tuning of protein expression of multigene operons, which is of biotechnological interest (96). As far as we are aware, four intergenic regulatory elements have so far been described in *B. subtilis* (see Table S1 in the supplemental material). The first one is the iron-responsive element between *feuA* and *feuB* mentioned above in the section on 3′UTRs (94). The second is situated between the *pstS* and *pstC* genes in the *pstSCABABB* operon (97) (too small to be annotated as an RNA segment in reference 15). The secondary structure between the *pstS* and *pstC* genes enables tuning of mRNA levels by an RNA processing event that most likely triggers degradation of the processed mRNA downstream of *pstS*. The third known intergenic region is the EAR switch, which is situated between the *epsB* and *epsC* genes. The EAR element is suggested to bind to RNA polymerase and thereby to allow for the processive antitermination of the unusually long transcript-encoding genes related to capsular polysaccharide biosynthesis (98). A fourth reported intergenic regulatory RNA is the GlmS ribozyme, which is situated between the *glmM* and *glmS* genes. This well-studied *B. subtilis* regulatory element integrates ligand binding and ribozyme activities (i.e., a ribozyme-riboswitch). It tightly regulates the expression of the *glmS* gene, which encodes an enzyme whose metabolic product is glucosamine-6-phosphate (GlcN6P). This gene is expressed at tightly regulated levels, and the *glmS* ribozyme-riboswitch takes care of this by activating cleavage and degradation of the RNA upon GlcN6P binding (29, 63).

The total number of RNA segments that were annotated as intergenic regions is 505, and this corresponds to 32% of all RNA segments. This relatively large proportion of the total number of RNA segments might point to many more additional regulatory mechanisms mediated by intergenic regions. Overview plots of the conservation, secondary structure, and length of the all-intergenic segments reveal a relationship between the GC content of each segment and its secondary structure score (Fig. 10). The all-intergenic category contains 21% asRNAs. What distinguishes these all-intergenic asRNAs is their length, since they are much longer than the all-intergenic sense RNAs. The all-intergenic asRNAs also have, on average, weaker secondary structures than the all-intergenic sense segments (Fig. 6). The three most conserved sense segments from the all-intergenic category are S56 (62/62 genomes), S928 (61/62 genomes), and S50 (57/62 genomes). S56 is situated between the highly conserved and essential genes *rpsK*, encoding ribosomal protein S11, and *rpoA*, encoding

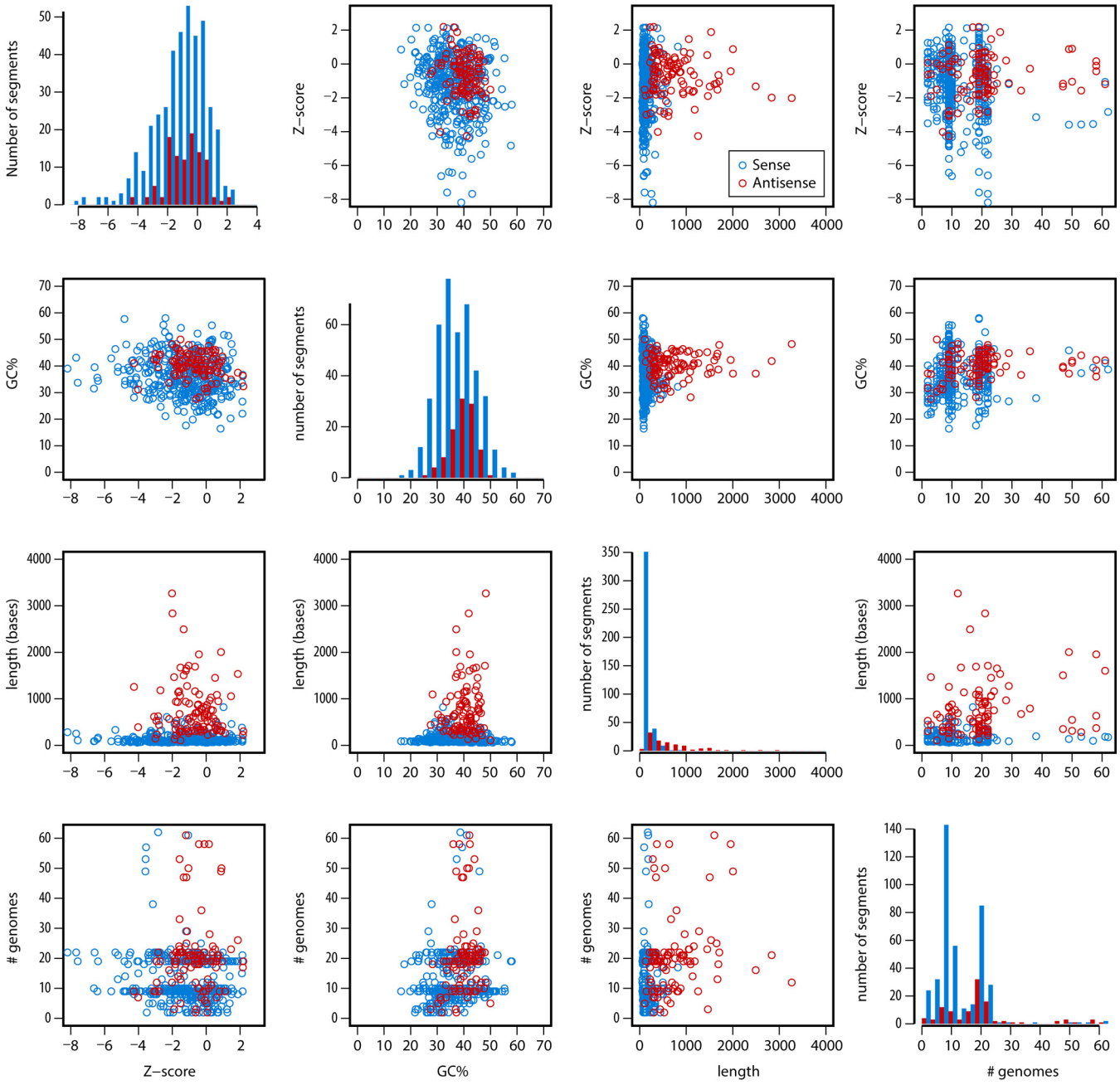

FIG 10 Overview plot of the predicted secondary structure MFE Z-scores, GC%, length, and species-level conservation for RNA segments of the all-intergenic category. For details on the calculation of MFE Z-scores and abbreviations, see the legend to Fig. 7.

the RNA polymerase alpha subunit. S928 is situated between *rsbRD*, encoding a probable stressosome component, and the *mgsR* gene, encoding a general stress-activated transcription factor. The antisense sequence of S928 is annotated as a tRNA-Gln; however, it was barely expressed under the thus-far-tested conditions (15). Despite this, the sequence similarity between S928 and tRNA-Gln may be the reason for the observed conservation of S928, even though S928 is larger than the tRNA-Gln (185 bases and 75 bases, respectively). S50 is situated between the highly conserved ribosomal protein-encoding genes *rplK* and *rplA*.

The 20% most structured (lowest MFE Z-score) and 20% most conserved segments from the all-intergenic category were selected and sorted based on their secondary structure MFE Z-scores (Table 1). Two of these segments, S1476 downstream of *efeB* (22/62 genomes, MFE Z-score −8.2) and S239 (22/62 genomes, MFE Z-score −7.7), have already been discussed in the previous section on 3′UTRs, since these were previously annotated as 3′UTRs (35). S1476 was annotated as an Inter segment by Nicolas et al. (15) because, under some conditions, transcriptional read-through seems to continue into the downstream segment. However, Rasmussen et al. noted that the gradual decrease in expression level is a characteristic of the class of 3′UTRs to which S1476 belongs (35).

Remarkably, S239 does not seem to be a 3′UTR but rather an intergenic region since it is transcribed between the essential *dagK/dgkB* gene, encoding a diacylglycerol kinase involved in lipoteichoic acid synthesis (99), and the *rlmCD/yefA* gene, encoding an rRNA methyltransferase responsible for 23S rRNA maturation. The condition-dependent expression profile of this region indeed seems to point to a regulatory role of S239, since the transcript levels of *dagK/dgkB* upstream of S239 are higher than those of *rlmCD/yefA* downstream of S239. Studying the importance of this type of RNA sequence element might unveil some of the remaining secrets of posttranscriptional regulation in *B. subtilis*.

Another intergenic RNA segment, S160 (22/62 genomes, MFE Z-score −5.4), contains the Bacillaceae-1 structural motif of unknown function (66). S160 is transcribed under some conditions from a gene that is located between *ydbT*, encoding a protein that confers resistance against antimicrobial compounds from *B. amyloliquefaciens*, and *acpS*, encoding an acyl-carrier protein synthase involved in fatty acid biosynthesis. S160 is predicted to fold into an intricate secondary structure, and *ydbT* and S160 are specifically induced under conditions of salt stress. Although S160 is probably intergenic, it cannot be excluded that it functions as an independent segment that arises from the *ydbT* promoter. The top-10 segments also include an 81-nucleotide-long intergenic RNA segment, S1554 (21/62 genomes, MFE Z-score −5.2), which is part of a tRNA operon. Since S1554 is highly structured and forms two stem-loops, it may regulate the transcription of these tRNA genes. Analogously, four of the 10 other highly structured and conserved intergenic RNA segments are ~90 bases in length and may function in regulating transcription or translation of their respective downstream genes. These are S565 (10/62 genomes, MFE Z-score −6.6) between *sigG*, encoding the forespore-specific late sigma factor, and *ylmA*, encoding a protein of unknown function; S600 (22/62 genomes, MFE Z-score −6.6), between *asd* and *dapG*, which are both involved in biosynthesis of lysine and peptidoglycan; S424 (13/62 genomes, MFE Z-score −5.2), between the sporulation-induced genes *yjcZ*, encoding a putative phage protein, and *spoVIF*, which is essential for spore coat assembly and resistance; and S1524 (19/62 genomes, MFE Z-score −4.8), between the myoinositol catabolism genes *iolH* and *iolI*. Based on its expression profile, the longer 164-nucleotide S918 segment (19/62 genomes, MFE Z-score −6.4), between the gene of unknown function *yqhY* and the *nusB* gene for a probable transcriptional termination regulator, may function to control read-through transcription between these two genes. Finally, the identification of the Inter segment S1503 (19/62 genomes, MFE Z-score −7.6) seems to relate to an incomplete genome annotation in the sense that it is transcribed between the purine riboswitch and the *yxjA/nupG* gene involved in purine uptake and is thus likely a structural part of this riboswitch.

Independently Expressed Small RNAs and RNA Antitoxins

RNA segments that are expressed from individual promoters and that are not attached to protein-encoding genes have been annotated as independent segments (15). These independent segments can have a wide variety of functions. They can, for instance, be *trans*-acting RNAs, here referred to as small regulatory RNAs (sRNAs), RNA antitoxins, or *cis*-acting asRNAs. As illustrated in the above sections, the automated annotation of RNA segments based on expression levels also made it possible to annotate RNA leader regions that control operon expression as Indep segments (e.g., the published *asnH* 5′ leader region S1534 [78]). To date, there are four functionally described sRNAs (FsrA/S512, SR1/*ykzW*, RsaE/S415/RoxS, and RnaC/S1022) and three RNA (type I) toxin-antitoxin systems in *B. subtilis* (*txpA/yqdB-ratA*/S976, *bsrG*/S809-SR4/S810, and *bsrE*/S717) (see Table S1 in the supplemental material). Functionally described here means that there is minimally one confirmed regulatory target reported for the sRNAs and that there is experimental proof for each of the two toxin-antitoxin systems.

The sRNA FsrA/S512 is an important regulator of the iron-sparing response of *B. subtilis* (43, 100). FsrA/S512 thus serves to reduce the expression of proteins relating to iron metabolism and storage. This role in the iron-sparing response makes FsrA/S512 functionally related to the *E. coli* sRNA RyhB. As such, this is an example of apparent functional conservation of an sRNA without sequence conservation (101). Three small basic proteins (FbpA, FbpB, and FbpC) have been shown to be involved in different aspects of this regulation, and there are experimental indications that at least FbpB may act as an RNA chaperone (100). Two sRNA-mRNA target interactions out of the most likely larger FsrA/S512 regulon have been experimentally confirmed, supporting strong indications from bioinformatics and expression analyses. First, the gel mobility of the *sdhC* mRNA, encoding succinate dehydrogenase, was shown to be reduced by addition of FsrA/S512, pointing to a direct *in vivo* interaction (43). Second, compensatory mutations were made in both *gltA*, encoding glutamate synthetase, and FsrA/S512. The individual mutations in either *gltA* or FsrA/S512 abolished the interaction while the presence of the respective complementary mutations restored the interaction (102). It is likely that the FsrA/S512 regulon extends to the expression of aconitase (encoded by *citB*), the lactate oxidases (encoded by *lutABC*), the C_4-dicarboxylate permease (encoded by *dctP*), the exocytoplasmic thioredoxin ResA, and the cytochrome bc_1 component QcrA (102). These targets can all be rationally linked to the iron-sparing response (102), making this the only documented sRNA regulon in *B. subtilis*. This implies that, as in Gram-negative bacteria (103), RNA regulons with multiple targets from the same or functionally related processes are also present in *B. subtilis* and thus in Gram-positive bacteria.

The SR1 segment (referred to by the original name of the ORF *ykzW*) was the first sRNA for which a function and target were reported in *B. subtilis*. SR1 inhibits the translation of *ahrC*, which encodes the transcriptional activator of the *rocABC* and *rocDEF* operons involved in arginine catabolism (104). The interaction between SR1 and *ahrC* is atypical in the sense that this sRNA does not bind to the 5′ end of the cognate mRNA. Instead, there are seven regions of complementarity between the coding part of *ahrC* and the 3′ end of SR1. SR1 has been reported to be exclusively expressed under gluconeogenic conditions, because of its repression by CcpN and CcpA (44). SR1 was later found to encode a 39-amino-acid peptide that increases mRNA stability of the *gapA* operon by binding to the GapA protein via an unknown mechanism. This made SR1 the first established dual-function sRNA in *B. subtilis* (105). Importantly, the two SR1 functions—regulation of *ahrC* mRNA via an sRNA-mRNA interaction and regulation of *gapA* mediated by the SR1 peptide—are conserved in related *Bacillus* species (106).

Illustrating its high conservation, RsaE/S415/RoxS was first identified in *Staphylococcus aureus* through a bioinformatics screen of this organism's intergenic regions (107). More recently,

the RsaE/S415/RoxS expression level was found to be induced by nitric oxide in a ResD-dependent manner. Consistent with a role in respiration via ResD, this sRNA is involved in the regulation of genes involved in *B. subtilis* redox homeostasis (108). Importantly, two direct RsaE/S415/RoxS targets were confirmed in this study. The first is the gene *ppnKB*, which encodes a kinase responsible for the production of $NADP^+$ from NAD^+. The second direct target is the central carbon metabolism gene *sucC*, which encodes the first subunit of the succinyl coenzyme A (CoA) synthase. Interestingly, RsaE/S415/RoxS requires processing for regulation of the *sucC* target (108).

RnaC/S1022 was first identified in a microarray screen of *B. subtilis* intergenic regions, and its expression was found to be completely dependent on the sigma factor SigD during the exponential growth phase on LB medium (24, 109). Recently, RnaC/S1022 was shown to modulate the cellular levels of the important transcriptional regulator AbrB via a direct sRNA-mRNA interaction (24). Interestingly, this repressive sRNA regulation of AbrB results in an increased cell-to-cell variation of the AbrB levels, as was demonstrated with green fluorescent protein (GFP)-tagged AbrB. Linked to AbrB being essential for *B. subtilis* growth in the employed M9 minimal medium, the RnaC/S1022-induced diversity in the cellular AbrB protein levels leads to heterogeneity in the growth rates of different cells within one population during the exponential phase (24). It was proposed that this sRNA-induced growth rate heterogeneity could reflect a bet-hedging strategy for enhanced survival under unfavorable conditions. In addition, the studies on RnaC/S1022 suggest that noise generation is a new and perhaps more general feature of sRNA regulation.

The *txpA-ratA* (*yqdB*/S976 in reference 15) toxin-antitoxin system is situated on the *skin* element of prophage origin. The two respective RNA species overlap by 120 nucleotides and interact via a so-called kissing loop interaction. This interaction between *txpA* mRNA and the *ratA* RNA promotes degradation of the duplex in an RNase III-dependent manner (53, 110, 111). In the absence of *ratA*, the toxic TxpA peptide can be translated, which causes cell lysis. The mechanism for TxpA-induced lysis is not known. However, its predicted N-terminal transmembrane domain and C-terminal charged residues point to a function in pore formation in the membrane. The physiological functions of toxin-antitoxin systems are still a matter of debate, but the *txpA-ratA* system conceivably helps maintaining the *skin* prophage in the *B. subtilis* chromosome (112).

Analogously to *txpA-ratA*, the *bsrG*-SR4 (S810-S809 in reference 15) toxin-antitoxin system is located on a prophage (i.e., SPβ) where *bsrG* encodes a toxic peptide of 38 amino acids. As is generally observed for toxin-antitoxin modules, the SR4 promoter is stronger (6- to 10-fold) than the *bsrG* promoter (113). Like *txpA-ratA*, the *bsrG* and SR4 RNA molecules interact at their 3′ ends. This interaction promotes degradation of the *bsrG* mRNA through cleavage of the duplex by RNase III and subsequent degradation of the remaining RNAs by endonuclease Y and the 3′-to-5′ exoribonuclease R (112). However, RNase III is not essential for the functioning of the toxin-antitoxin system (113), and this might be due to a second role of SR4 in inhibiting *bsrG* mRNA translation (114). As for the *txpA-ratA* module, it was suggested that the function of the *bsrG*-SR4 module is to prevent excision of the SPβ prophage element from the *B. subtilis* genome. In this respect, it is important that, similarly to the situation in other well-characterized type I toxin-antitoxin systems, the BsrG toxin-encoding mRNA is long-lived with an identified half-life of ~15 min, while the SR4 antitoxin RNA is relatively short-lived with a half-life of ~3 min. This difference in RNA half-lives is the basis for the functioning of the toxin-antitoxin system, because if the SPβ prophage region is lost from the genome, SR4 is degraded relatively quickly, while *bsrG* remains present much longer. The resulting uninhibited translation of the toxin from the *bsrG* gene will then kill the cell that has lost the SPβ prophage. The functioning of *bsrG*-SR4 may be linked to temperature stress, since the *bsrG* mRNA was destabilized 3.5-fold by heat shock at 48°C while the stability of SR4 was found to remain unaffected under this condition (113).

The *bsrE*/S717 toxin-antitoxin system is located on the P6 prophage (115). The *bsrE*/S717 RNA segment encodes a peptide of 30 amino acids, and its antitoxin is SR5/S718. Consistent with the toxin-antitoxin model, the condition-dependent expression of *bsrE*/S717 is much more plastic than that of its antitoxin SR5/S718 (15). It was recently found that overexpression of *bsrE*/S717 leads to cell lysis on agar plates and that the two RNA molecules form a perfect duplex, the formation of which is possibly mediated by a thus-far-unknown RNA chaperone. In a series of detailed experiments, multiple loop regions of SR5/S718 were found to be essential for triggering RNase III-mediated *bsrE*/S717 degradation (115).

Overview plots of the conservation, secondary structures, and lengths of the 153 RNA segments from the all-independent category reveal a trend toward stronger predicted secondary structure for both more conserved segments and shorter segments (Fig. 11). The majority of the longer segments are asRNAs, and there is actually a significant difference between the MFE Z-score of the antisense and the sense all-independent segments (Fig. 6). In fact, more than half (58%) of the all-independent segments are asRNAs. The most conserved sense segments of the all-independent group are the RsaE/S415/RoxS sRNA (57/62 genomes), the leader region S198 of the *vlmR*-encoded ABC transporter (48/62 genomes), and S1455 (43/62 genomes) upstream of the threonyl-tRNA synthetase gene, which is responsible for an as-yet-unannotated T-box regulatory mechanism (see Table S1 in the supplemental material).

The 20% most structured (lowest MFE Z-scores) and 20% most conserved segments from the all-independent category were selected and sorted based on their predicted secondary structure MFE Z-scores (Table 1). Two of these 10 segments, as-*brsH*/S977 (18/62 genomes, MFE Z-score −7.5) and *bsrE*/S717 (17/62 genomes, MFE Z-score −4.9), are related to type I toxin-antitoxin systems. as-*brsH*/S977 is the putative antitoxin of *brsH*/S978. This system was first described by Saito et al. (12), and it is located in the same intergenic region as *txpA-ratA* on the *skin* element. Since *bsrH*/S978 was found to have a 3′ overlapping asRNA, namely, as-*bsrH*/S977 (14), and putatively encodes a peptide, *bsrH*/S978 and *bsrH*/S977 were proposed to form a type I toxin-antitoxin system (reviewed in reference 112). The peptide encoded by *bsrH*/S978 is 29 amino acids in length, and it contains a single α-helical transmembrane domain and charged residues at the C terminus. Since this predicted peptide is very similar to that of *bsrE*/S717 (14, 112), this arrangement is consistent with that of other reported type I toxins. In addition, the as-*bsrH*/S977 expression level is similar across conditions (15), which is expected for an antitoxin. The definition of the exact mode of action of as-BsrH/S977 awaits further analysis, especially since it seems to act somewhat differ-

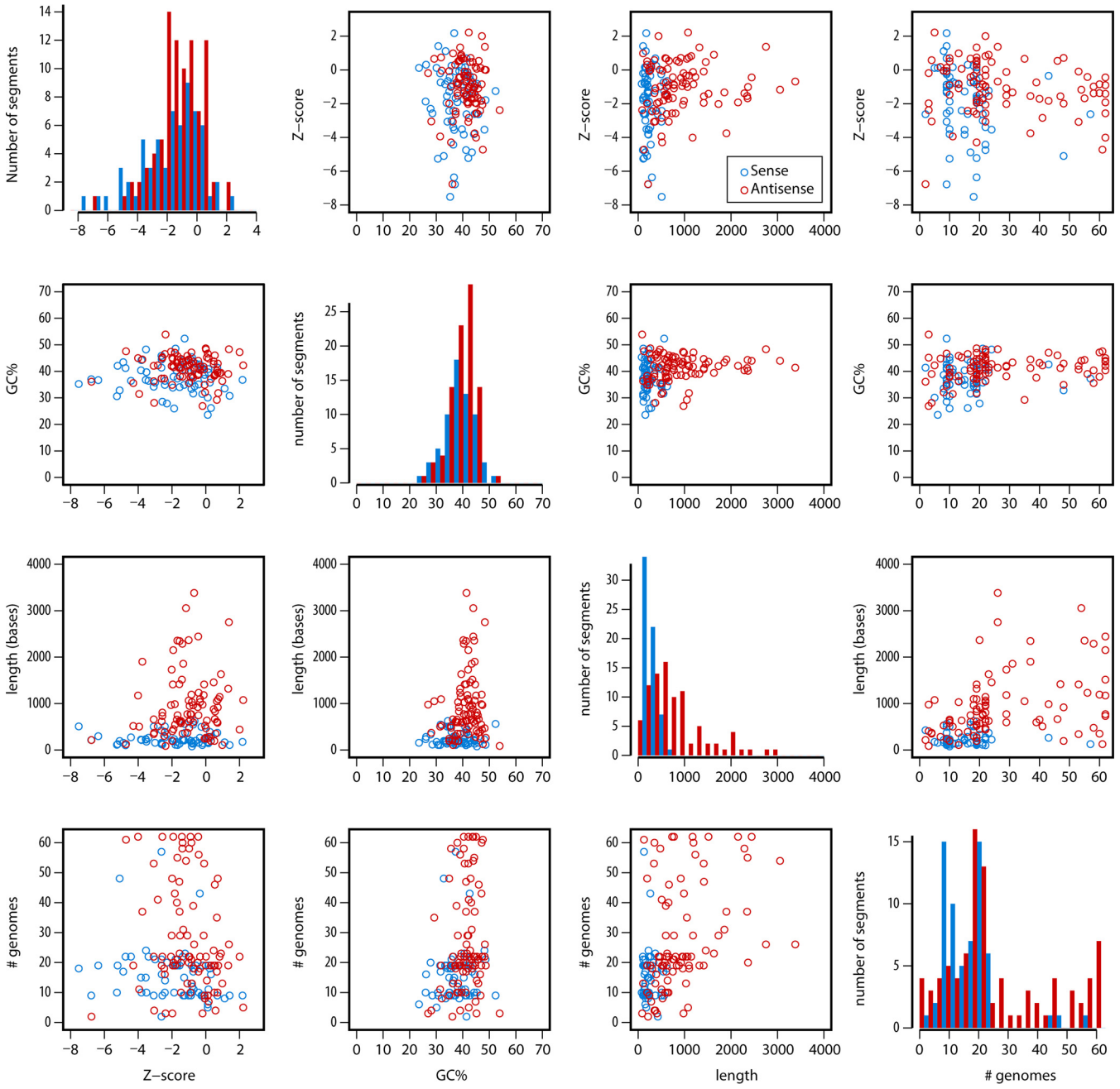

FIG 11 Overview plot of the predicted secondary structure MFE Z-scores, GC%, length, and species-level conservation for RNA segments of the all-independent category. For details on the calculation of MFE Z-scores and abbreviations, see the legend to Fig. 7.

ently than the currently described type I toxin-antitoxins in *B. subtilis* (112). The next top-10 segment in the all-independent category is S198 (48/62 genomes, MFE Z-score −5.1), which forms the leader region of the *vlmR*-encoded ABC transporter that was briefly addressed above. Transcription termination-mediated control of read-through transcription into *vlmR* has been reported and seems to be mediated by S198 (58). However, this function does not necessarily exclude a second regulatory mechanism for S198 as an sRNA, especially since such a dual-function RNA has been reported in *Listeria monocytogenes* (116).

Three other segments from the top-10 list, namely, S2 (22/62 genomes, MFE Z-score −4.4), S326 (10/62 genomes, MFE Z-score −5.3), and S1292 (19/62 genomes, MFE Z-score −4.3), are transcribed antisense and thus may have a regulatory function in *cis*. It should be noted, however, that S2 is not annotated as an asRNA, since it does not overlap a protein-encoding gene but overlaps the 5′UTR region (S1 and S3) of *dnaA*, which encodes the replication initiation protein. The expression level of S2 is relatively low but most prominent under conditions of sporulation initiation. S326 is transcribed from the opposite strand of the intergenic region of the *nsrR* and *ygxB* genes. As was the case for S2, for S326 the highest expression level was also observed under sporulation con-

ditions. Judged by the functional annotation of its neighboring genes, S326 could perhaps be involved in the repression of ResD-ResE-dependent genes in the absence of nitric oxide (similarly to *nsrR*) and/or the general stress response (similarly to *ygxB*). S1292 is expressed at a relatively low level, and this RNA segment is transcribed antisense to the *yvaV* gene, encoding a protein similar to the transcription factor OpcR, and the *sdpA* gene, encoding a protein that is required for SdpC toxin maturation.

Three previously reported sRNAs are also part of the top-10 list, namely, FsrA/S512 (22/62 genomes, MFE Z-score −4.8), RnaC/S1022 (19/62 genomes, MFE Z-score −5.2), and CsfG/S547 (61/62 genomes, MFE Z-score −4.7). As discussed above, FsrA/S512 is an important regulator of the iron-sparing response of *B. subtilis* (43, 102). Interestingly, in the tiling array study by Nicolas et al. (15), FsrA/S512 was found to be expressed also under conditions with plentiful iron. It therefore remains to be seen whether this reflects on possible additional FsrA/S512 functions besides those reported in the iron-sparing response. The function of RnaC/S1022 in regulating *abrB* (24) was already discussed above. The sequence on the opposite strand of CsfG/S547 was first annotated as a *cis*-acting regulatory element, referred to as the "*ylbH* leader" by Barrick et al. (62). However, Marchais et al. (117) have provided convincing evidence that the region on the opposite strand more likely encodes an sRNA, and they also noted its high conservation among endospore formers. The identification of the "*ylbH* leader" thus seems to be an artifact caused by the antisense location of CsfG/S547. In addition, Marchais et al. found that the sporulation-specific expression of CsfG/S547 is due to regulation by SigG and SigF (117). Importantly, their transcriptional fusion data are in accordance with the unsupervised promoter cluster analysis by Nicolas et al. (15). Interestingly, the final segment of the top 10 of the all-independent segments, S357 (19/62 genomes, MFE Z-score −6.4), seems related to CsfG/S547 at the expression level. First, S357 seems to be an sRNA, which is specifically induced under sporulation, fermentation, and anaerobic conditions. Second, promoter cluster analysis also predicts a sporulation sigma factor control of S357 expression. No additional regulators are predicted by the DBTBS database of transcriptional regulation in *B. subtilis* (http://dbtbs.hgc.jp) in the upstream region of S357 (data not shown). Notably, expression of the genes situated next to S357, *hit* and *escA*, is not correlated with S357 expression, and this suggests a completely independent *trans*-acting role for S357.

Antisense RNAs

RNA segments that overlap protein-encoding genes with more than 100 bases or >50% of their sequence due to transcription from the opposite DNA strand are here referred to as asRNAs. While it is doubtful whether all asRNAs have a role in *cis*, it is clear that different asRNAs can regulate their sense genes via a wide variety of mechanisms. In this respect, one has to bear in mind that the transcription of an asRNA does not automatically lead to regulation due to the structural complexity of RNA, where even completely complementary sequences could fold in such a way that no duplex formation is allowed. A recent review on *cis*-acting asRNAs in bacteria (118) distinguished four main mechanisms of asRNA regulation including (i) alteration of target RNA stability, (ii) modulation of translation, (iii) transcription termination, and (iv) transcriptional interference. However, subcategories of these regulatory mechanisms can be identified as well (119, 120). All these types of regulation can alter the sense mRNA levels, either directly or by preventing ribosome protection of the mRNA. When the focus is on negative regulation, this implies that an anticorrelation with its putative sense RNA can be regarded as a clue for the functionality of an asRNA.

An asRNA-RNA interaction is thought to take place via two binding mechanisms, which have been identified through research on plasmid- and transposon-borne asRNAs, namely, one-step or multistep binding (119). The one-step binding mechanism is exemplified by the RNA-IN/RNA-OUT system, which regulates the transposition of IS*10* in *Escherichia coli*. Here, the contact between sense and antisense is made directly via the interaction of the stem-loop of one RNA molecule with the single-stranded region of the other RNA molecule (reviewed in reference 119). This initial contact is then extended to form a full duplex. In the multistep binding mechanism, additional steps are required for the formation of a full duplex. For example, the CopT-CopA interaction, which regulates the replication of *E. coli* plasmid R1, starts with a reversible interaction between the stem-loops of both the sense and antisense RNAs. This interaction is fixed in a so-called kissing complex, which subsequently extends to result in a four-helix junction and finally a stable inhibitory complex (reviewed in reference 119). Such asRNA-RNA duplexes are eventually degraded by RNase III in *E. coli*. Notably, however, the formation of duplexes over the complete length of the molecule is often not required for regulation (119), and limited sequence interactions can thus be sufficient for asRNA regulation.

There are seven confirmed sense-antisense interactions in *B. subtilis*. The first three are represented by the type I toxin-antitoxin systems *txpA-ratA*, *bsrG*-SR4, and *bsrE*/S717, as discussed in the previous section. The fourth is an asRNA named S25 (15), which is under the control of the extracytoplasmic function (ECF) sigma factors SigM and SigX (121). S25 is expressed antisense to the *yabE* gene, which is predicted to encode an autolysin of the Rpf/Sps family implicated in stationary-phase survival and resuscitation of dormant cells. This antisense regulation was proposed based on a negative correlation between the two RNA species analyzed via Northern blotting (121). Indeed, the correlation of S25 and *yabE* under all expression conditions investigated by Nicolas et al. is also negative (−0.28). The fifth sense-antisense pair is *gdpP*/*yybt-gdpP*$_{as}$/S1559 (122). GdpP is a cyclic di-AMP phosphodiesterase responsible for cyclic di-AMP degradation. The presence of *gdpP*$_{as}$/S1559 was found to reduce the GdpP protein level in the cell, but no further phenotype in the absence of *gdpP*$_{as}$/S1559 was identified. Transcription of *gdpP*$_{as}$/S1559 is regulated by the alternative sigma factor SigD (122). The transcripts of *gdpP* and *gdpP*$_{as}$/S1559 display a small positive correlation of 0.15 under all expression conditions reported by Nicolas et al. (15). In addition, there was no obvious induction of *gdpP*$_{as}$/S1559 and concomitant downregulation of GdpP observed under any of the probed environmental conditions (15). Since an effect on the protein level was reported, this may imply that *gdpP*$_{as}$/S1559 regulates the *gdpP* mRNA at the level of translation initiation. The sixth sense-antisense pair, *cwlO*-S1326 (80), was found to be strongly negatively correlated (−0.62) (15). CwlO is a D,L-endopeptidase-type autolysin encoded by a transcript that is highly unstable due to RNase Y-dependent cleavage in its S1327 leader region, as mentioned above in the section on 5′UTRs (80). Transcription of the S1326 asRNA was found to depend on the general stress sigma factor SigB, and absence of the respective promoter caused cells to

enter exponential growth earlier than the parental strain. Since this phenotype was rather weak, it was suggested that S1326 might play a more important role under specific environmental conditions (80). The conditions under which S1326 displays its highest expression suggest that this asRNA may be most important under conditions of swarming (Sw), biofilm formation (B60), and the late growth phase on minimal medium (M9-stat) (15).

The seventh known sense-antisense pair is represented by *rpsD*-S1136/S1134 (70). *rpsD* is an essential gene encoding the S4 component of the small ribosomal subunit. The S1136 segment is transcribed from an independent convergent promoter on the opposite strand of *rpsD*. Of note, the transcription of S1136 is extended further downstream of S1136 due to incomplete termination, resulting in the S1134 segment. While S1136 has only a 23-nucleotide overlap with *rpsD*, S1134 covers the entire length of *rpsD*. It had previously been noted that many genes encoding components of the transcriptional and translational apparatus of *B. subtilis* are downregulated upon ethanol stress. The S1136/S1134 asRNA was first reported in the study by Nicolas et al., where it was found to be almost exclusively induced under the condition of ethanol stress. Concomitantly, *rpsD* was strongly downregulated upon ethanol stress, resulting in a strongly negative global expression correlation between S1136/S1134 and *rpsD* of −0.59 (15). The subsequent analyses revealed that this *rpsD* downregulation is due to the SigB-dependent induction of the S1136 promoter. Conceivably, this downregulation could take place via a promoter collision mechanism. In addition, it was found that the S1136 promoter-dependent *rpsD* downregulation results in a reduced abundance of the small ribosomal subunit upon ethanol stress (70). The observed reduction in the level of the small ribosomal subunit, which contains the ribosome-decoding center, may protect *B. subtilis* cells against misreading and spurious translation of possibly toxic aberrant peptides in the presence of ethanol.

In the study by Nicolas et al., 13% of the protein-encoding genes of *B. subtilis* were found to be overlapped by an RNA segment from the opposite strand (15). This proportion of potential asRNAs was similar to that found in a previous transcriptome study on *Mycoplasma pneumoniae* (123). Three important observations were made on the asRNA genes of *B. subtilis* in comparison to its protein-encoding genes (15). First, a low proportion of the *B. subtilis* asRNAs are regulated by the household sigma factor SigA. Instead, most asRNAs are regulated by alternative sigma factors. Second, many of the asRNA segments arise from imperfect control over transcriptional termination by read-through transcription from the 3′ ends of genes or operons. The third observation was that the conservation of promoter sequences controlling transcription of independently expressed asRNAs was just slightly above that of a random sequence. Therefore, Nicolas et al. proposed that many asRNAs arise in *B. subtilis* through spurious initiation from alternative promoter sequences and incomplete control over transcription termination (15). Interestingly, a recent study by Mader et al. (33) uncovered a relatively low abundance of asRNAs in *S. aureus* compared to *B. subtilis*. In *S. aureus*, asRNAs were found to overlap only 6% of the protein-encoding genes, and just 19 asRNAs were found that are not cotranscribed with other genes. Promoter analyses suggest that the small number of asRNAs in *S. aureus* can be related to the fact that there is usually only one of three alternative sigma factors, namely, SigB, active in *S. aureus*. At the same time, antisense transcription in *S. aureus* (and *B. subtilis*) is effectively suppressed by Rho-dependent transcription termination, potentially setting a limit to spurious transcription initiation. Of note, the results obtained for mapping the regulatory RNA segments of *S. aureus* (33) are directly comparable to those obtained for *B. subtilis* (15) as the same tiling array approach and highly standardized protocols were applied.

Although many asRNAs of *B. subtilis* may be derived from spurious initiation or ineffective termination, this does not exclude the possibility that some of the identified asRNAs have a function. For example, the expression of sense-antisense pairs was found to be significantly anticorrelated more often than random pairs of genes. On the other hand, this observation alone is not indicative of an overall functional role of the identified asRNAs, as it could be due to sigma factor bias in asRNA regulation or result from purifying selection that eliminates the asRNAs that impact expression of the sense strand. Overall, this means that the newly identified asRNAs have to be analyzed on a case-by-case basis for possible functions. It is also important to search for phenotypes other than downregulation of the opposite gene since this may potentially arise as a simple by-product of antisense transcription. Ideally, a phenotype that could convincingly contribute to fitness should be identified, as was exemplified for *rpsD*-S1136/S1134 (70).

To identify putative functional asRNAs, we first sorted all sense-antisense pairs reported by Nicolas et al. (15) based on their negative expression correlation, and the two pairs with the strongest anticorrelation are discussed in the following paragraph. Second, we addressed four independently expressed asRNAs which display a negative expression correlation and are regulated by the stress sigma factor SigB, since these may be part of this well-studied general pathway for stress adaptation (124).

The transcriptionally most negatively correlated asRNA is S31, which is independently transcribed from a convergent promoter antisense to the conserved *ispE-purR-yabJ* operon. The S31 promoter is predicted to depend on early sporulation sigma factors (SigE and SigF), and the expression correlation of S31 with the *yabJ* gene is −0.83. The YabJ protein is a transcription factor required for activity of its neighboring gene *purR*, which is involved in the regulation of purine biosynthesis. This observation is interesting, since PurR represses purine biosynthesis (125) and purine nucleotides can induce sporulation (126). In addition, S31 is induced at the onset stage of sporulation, before the final commitment to sporulation is made (127). We therefore propose that S31 may be part of a last-effort response of the cells to increase their levels of purine nucleotides by repressing the repressor of purine biosynthesis. The second asRNA segment with a strong negative transcriptional correlation is S526, which is also dependent on the early sporulation sigma factors. S526 is transcribed antisense of the two-component sensor kinase *kinC*. Spo0A has been reported to transcriptionally activate the SigA-dependent *kinC* promoter at the onset of sporulation, while expression was unexplainably observed to shut down at a later time point (128). This expression pattern seems fully consistent with the expression data, where *kinC* was found to be expressed in the first hours of sporulation but repressed soon thereafter at a time point that seems to correspond exactly with the time point of S526 induction (15). This results in a strong negative transcriptional correlation of −0.76, which suggests that antisense regulation via S526 might be responsible for shutting down *kinC* expression.

Besides these two examples of likely functional sense-antisense pairs, we subsequently looked at those sense-antisense pairs where

the asRNA is under the (predicted) control of the stress sigma factor SigB and sorted these again based on their transcriptional correlation. Interestingly, the *cwlO*-S1326 pair (80) exhibited the highest negative correlation (−0.62), and this pair was already described above. The second most negative correlation was observed for the *yugH*/*alaT*-S1202 pair (−0.46). Since *yugH*/*alaT* encodes a protein of unknown function and expression of S1202 is low under all conditions, we did not consider this pair any further. A third pair with a strongly negative transcriptional correlation is *dacA*-S9 (−0.39). DacA is a D-alanyl–D-alanine carboxypeptidase, also known as penicillin-binding protein 5*, which is involved in the maintenance of the cell wall peptidoglycan. Consistent with the SigB-dependent regulation of S9, this asRNA is specifically induced under stressful conditions, most prominently under conditions of ethanol stress. In turn, *dacA* expression is specifically reduced under ethanol stress. The relevance of *dacA* downregulation might be related to the effect of ethanol on the peptidoglycan cross-linking (129), and it seems plausible that this function is provided by the S9 asRNA. The fourth considered SigB-dependent asRNA is S931, which is transcribed antisense of the *yqgS* gene, encoding a polyglycerolphosphate lipoteichoic acid synthase involved in the biosynthesis of lipoteichoic acid. Lipoteichoic acid is an amphipathic polymer found in the cell wall of many Gram-positive bacteria (130). The negative correlation between S931 and *yqgS* of −0.37 is a consequence of the strong ethanol stress induction of S931 and the concomitant downregulation of *yqgS*. A direct role for lipoteichoic acids in stress adaptation has not been reported. However, since ethanol has a strong effect on the cell envelope in general, it is conceivable that this is a so-far-unappreciated aspect of the SigB response.

RNA Chaperones in *B. subtilis*

RNA regulation may be mediated by chaperone proteins. In many bacteria, Hfq is the main RNA chaperone. Hfq is a hexameric protein with a tertiary donut-like conformation that mediates RNA-RNA interactions in ways that are not yet completely understood (for a review, see reference 131). Discussions about regulatory RNAs in bacteria have become tightly linked to the function of Hfq. However, this seems mostly due to the importance of Hfq for RNA-based regulation in two closely related Gram-negative bacteria, *Escherichia coli* and *Salmonella*, which both belong to the *Enterobacteriaceae*. In both these model organisms, Hfq is indispensable for much of the identified RNA regulation. Despite this, genes encoding Hfq homologues are found only in approximately half of the currently sequenced genomes (22), indicating that there may be alternative ways of mediating RNA regulation. This view is underscored by studies on *B. subtilis*, which contains an Hfq homologue (previously named YmaH) that is dispensable for the sRNA-mRNA interactions reported in this organism so far (23, 24). In addition, recent studies on the influence of *B. subtilis* Hfq at the transcriptome level have shown that while Hfq has some influence on RNA regulatory processes, this was not to the extent that has been reported for other species (23, 132). Since Hfq in Gram-negative bacteria interacts with both the sRNAs and mRNAs, precipitation of Hfq-interacting RNA followed by deep sequencing has greatly advanced our current understanding of regulatory RNAs and their targets in these bacteria.

It is presently unclear whether the lack of strong phenotypes for the *B. subtilis hfq* mutant means that there is another RNA chaperone protein central to RNA regulation in *B. subtilis* or that there is perhaps no requirement at all for such a chaperone. Notably, it has been suggested that the requirement of an RNA chaperone may be linked to the structure, size, and GC content of the species genome (22, 133). For instance, some sRNA-mRNA interactions in *S. aureus* are reported to be generally more stable than those in *E. coli*, which could abolish the requirement for a chaperone (22, 133). However, this might be linked to the specific example of *S. aureus* RNAIII. It may also be that for RNA interactions in *B. subtilis* that do require a chaperone, a dedicated condition-specific RNA chaperone has evolved, which would limit the dependence on a global RNA chaperone mediating the vast majority of all RNA-RNA interactions. A condition-dependent RNA chaperone model seems to be suggested by investigations on FsrA, which to date is the most comprehensively studied *B. subtilis* sRNA. The regulation by FsrA is linked to the presence of three small basic proteins, and together they mediate the *B. subtilis* iron-sparing response (43, 100, 102). These three small basic proteins could thus (in part) function as RNA chaperones (100). More recently, it was shown in *L. monocytogenes* that the widely conserved SpoVG protein is in fact an RNA-binding protein that can interact with noncoding regulatory RNAs (134). In *B. subtilis*, SpoVG has been implicated in asymmetric cell division and sporulation but is highly expressed under many conditions. This could indicate that SpoVG has a general RNA chaperone function. Of note, the finding that Hfq seems of relatively limited importance for RNA regulatory processes in *B. subtilis*, at least under laboratory growth conditions, cannot be directly extrapolated to other bacilli. For instance, sequencing analyses have shown that *Bacillus anthracis* harbors three Hfq homologues, suggesting a selective advantage of Hfq during the evolution of this species (135).

Evolutionary Aspects of Regulatory RNA Regulation

The evolutionary conservation of genes is an indication of their biological importance. In this respect, it is noteworthy that the repertoire of regulatory RNAs is highly variable between distant species. The generally low level of conservation of regulatory RNAs is also reflected in the conservation analysis presented here. Only a small set of the identified RNA segments was found to be conserved in all the analyzed *Bacillus* genomes (Fig. 3). In fact, most of these segments are conserved only in the genomes of *B. subtilis* and *B. amyloliquefaciens* isolates. This low level of sequence conservation is consistent with the observation that *cis*-acting regulatory elements of highly conserved genes are often highly diverged between species (101, 136). Because of this large variability in RNA regulators between species, it may in some contexts be more relevant to speak about the conservation of regulatory RNA functions than of sequence conservation. A good example of such functional conservation is provided by the *E. coli* RyhB and *B. subtilis* FsrA/S512 RNA segments in the iron-sparing response (101).

One intriguing question that should be asked in this context is why regulatory RNAs are so variable between species. Possible answers to this question may follow two lines of reasoning. The first relates to the evolution of particular regulatory RNA sequences, and the second relates to the evolution and conservation of particular regulatory functions. In this context, one should bear in mind that slight sequence alterations (e.g., 1 or 2 bp) can be sufficient for the creation of a new promoter sequence and the transcription of a new regulatory RNA. Furthermore, such new RNAs can be made with low energetic costs compared to proteins,

because the energy-intensive translation step is not required (137). A high probability of the mutagenesis-driven emergence of new promoter motifs or the disabling of a terminator structure was also proposed by Nicolas et al. in arguing that many of the newly identified RNA segments in *B. subtilis* could be spuriously transcribed and may therefore have no function (15). However, such randomly emerging segments can represent a pool for the evolution of new regulatory capacities that can be employed to control new biological functions. As such, they would contribute to the "evolvability" of an organism, meaning its capacity to generate heritable phenotypic variation (138). These new functions can be implemented either in the form of base-pairing regulatory RNAs, asRNAs, or, if they contain an ORF, as a source of new protein-encoding genes (139), or combinations thereof. It is thus conceivable that many of the RNA segments of *B. subtilis* may have originated spuriously but have now adopted subtle, condition-specific regulatory functions. This would be facilitated by the initiation of transcription by alternative sigma factors that readily confer condition-specific expression.

Condition dependency has large consequences for the evolution and conservation of a regulatory function. Notably, this is true for protein-encoding genes as well as regulatory RNA genes. However, since the latter directly rely on the nucleotide sequence without a requirement for meaningful translation (i.e., RNA-RNA interactions do not involve a wobble base), the evolution of new regulatory functions will potentially occur more frequently for regulatory RNAs. Another driving force for quickly evolving RNA regulation can be the modularity of regulatory RNA molecules. This means that chimeric regulatory RNA molecules can easily originate from recombination events. In addition, regulatory RNAs can be prone to gene duplications, as was reported by Rasmussen et al. (35) and in the above section on 3′UTRs. One intriguing evolutionary question is whether a regulatory function of a regulatory RNA has to remain conserved when a regulatory RNA sequence is conserved. Indeed, it seems that this is not always the case (101). However, it would still be very interesting to test whether, for instance, RsaE in *S. aureus* has retained a shared regulatory function with RsaE/S415/RoxS in *B. subtilis*. Clearly, due to the speed of regulatory RNA evolution, the expression conditions and regulatory functions of RsaE could be very different in the two species. Alternatively, it is possible that one function is still shared between the two while other functions have been coopted. This is a question that needs to be addressed in future studies. For a more general discussion on evolutionary aspects of regulatory RNA regulation, please see the recent review by Updegrove et al. (101).

CONCLUSION AND OUTLOOK

With the identification of hundreds of putative regulatory RNA segments in a multitude of bacteria, it has become a major challenge to analyze and dissect their functions. The recently published rich condition-dependent transcriptomics data set for *B. subtilis* not only delivered 1,583 new RNA segments but also provided the means to predict functions for these RNA segments. The goal of the present review was to provide an inventory of reported regulatory RNAs in *B. subtilis* and to classify the identified RNA segments by their level of evolutionary conservation, predicted secondary structures, and expression levels. For all categories, the top-10 lists of most conserved and structured RNA segments (Table 1) were discussed in some detail and, whenever possible, placed in a biological context to predict their functions. It now awaits further functional analyses to find out the extent to which these and other predictions that could be drawn from the material provided in Tables S2 and S3 in the supplemental material are correct. In the subsequent last paragraphs of this review, a few additional considerations for the future study of regulatory RNAs will be discussed with a special focus again on *B. subtilis*.

Functional Regulatory RNAs in *B. subtilis*: Where and How Many?

Because of the limited number of functionally studied *B. subtilis* regulatory RNAs, it seems premature to make far-reaching general statements about the preferentiality of RNA-based regulatory mechanisms. However, the following six observations can be delineated from the current state of the art as reviewed here. (i) For many *B. subtilis* genes, posttranscriptional regulation takes place at the level of transcriptional read-through, via metabolite-, tRNA-, or protein-controlled switches or other attenuation mechanisms. (ii) The 5′ end of RNA molecules seems to be the main determinant of RNA stability in *B. subtilis*. (iii) Hfq is not the central RNA chaperone of *B. subtilis*, as it is in many Gram-negative bacteria, but there might be other central or condition-dependent RNA chaperones. (iv) Similar to other bacteria, type I toxin-antitoxin systems seem important for maintaining prophage elements in the *B. subtilis* genome. (v) asRNAs that overlap protein-encoding genes are dominant and have been reported to affect the levels of their cognate sense mRNAs. Since the regulatory effects of asRNAs are local, it is conceivable that most functional asRNAs have subtle fine-tuning functions, which can be revealed only by carefully designed studies. In addition, many asRNAs probably arise as by-products from incomplete control over transcriptional termination. (vi) *trans*-acting sRNAs are present in *B. subtilis*. These sRNAs can mediate important physiological adaptations, and they can form sRNA regulons as exemplified by FsrA/S512. However, it may still be that sRNA regulons are a less common phenomenon in *B. subtilis* than in the Gram-negative bacteria *E. coli* and *Salmonella*. If so, this might relate to the presence of the large number of alternative sigma factors in *B. subtilis*. Notably, *B. subtilis* has 17 reported sigma factors, of which 4 are sporulation specific. This number of sigma factors is substantially larger than the number observed in *E. coli*, which has 7 reported sigma factors. Thus, it is conceivable that alternative sigma factors mediate coordinated responses in *B. subtilis* while similar responses could be regulated by sRNAs in species with fewer sigma factors. It will thus be very interesting to assess whether there might be an inverse relationship between the number of sigma factors and the importance of sRNA regulation in different bacterial species. Importantly, this possibility can be addressed by studies on sRNA-mediated regulation in *S. aureus*, which is related to *B. subtilis* but has a far lower number of alternative sigma factors, a trait that was used by Mader et al. (33) to confirm the idea that the relatively high abundance of asRNAs in *B. subtilis* could be correlated with this bacterium's relatively high number of alternative sigma factors.

Nomenclature

In many cases where a function has been described for protein-encoding genes or regulatory RNAs, these have been renamed. Unfortunately, this has resulted in multiple names for the same gene, which is a frequent source of confusion. While we can agree

that renaming RNA segments may be useful in certain instances, we believe that this should be done in such a way that their original names remain traceable. For this purpose, it would be an attractive option to keep using the automated S-number annotation as generated by Nicolas et al. (15). Specifically, we suggest that it would be helpful to maintain the S-number of an RNA segment in association with a new name, like we have presently done for FsrA/S512, RnaC/S1022, and various other already-characterized RNA segments. We believe that this will greatly facilitate the traceability of data and comparisons with the earlier literature.

Outlook on the RNA Interactome of *B. subtilis*

The number of putative regulatory RNAs in *B. subtilis* is large, and accordingly, the RNA interactome is even larger. This seems to suggest that there are specific aspects in regulatory RNA regulation that are somehow preferable to protein-based regulation. Indeed, a modeling study has provided quantitative evidence that regulation by sRNAs is advantageous when fast responses to external signals are needed (140), which would be consistent with experimental data on the involvement of sRNA in bacterial stress management. The faster response time in RNA-mediated regulation relates to the omission of a translation step prior to the actual regulation step, since the regulatory RNA will be directly active upon transcription. Another advantage may be that the half-lives of RNA molecules are generally shorter than those of proteins, which makes the regulatory signals provided by RNA molecules intrinsically transient. This allows for an almost immediate change in regulatory output once the transcription of the respective regulatory RNA is deactivated. Such fast dynamics of RNA regulation also hold for antisense regulation, but importantly, asRNA-mediated regulation has an additional advantage, namely, that of signal integration. Transcription of a gene can be regulated by multiple factors, generally involving the binding of these factors in the form of proteins in the vicinity of the promoter region. This sequence-dependent regulation puts a constraint on the number of protein factors that can influence the transcript levels. The transcription of a functional asRNA controlled by an additional transcription factor expands the number of factors that can regulate the sense gene. Consequently, this ameliorates the constraints on the number of factors affecting transcript levels. It is noteworthy that this flexible tunability and the fast dynamics, together with the modularity of regulatory RNAs and the relative ease of sequence-based design, also make regulatory RNAs attractive for use in engineered regulatory circuits that can be implemented in (large-scale) synthetic biological systems (60). Indeed, synthetic regulatory RNA molecules have already been employed in bacteria to implement artificially designed regulatory systems with sensing, regulatory, enzymatic, and scaffolding functions (see references in reference 60).

The extensive set of putative regulatory RNA segments in *B. subtilis*, first identified by Nicolas et al. (15) and further explored in this review, opens up many possibilities for in-depth follow-up studies (24, 50, 70, 141). This can of course be approached through targeted studies in which the RNA-mediated regulation of a particular gene or regulon is assessed. However, because of the large number of putative regulatory RNA segments, their function is ideally approached in a high-throughput, or at least medium-throughput, manner. One can, for instance, envision a plasmid system in which all 5′UTRs can be cloned upstream of a reporter gene, and the resulting library could be assayed under different growth conditions to delineate the influence of the respective cloned 5′UTR on the expression of the downstream reporter gene. From such a screen, 5′UTRs of interest can be selected and studied in their natural context. Similarly, this could be done with cloning intergenic regions between two reporter genes and optionally with 3′UTRs by cloning them downstream of a reporter gene. In addition, a large suite of sequencing-based technologies have been developed for studying sRNA function (reviewed in reference 142), and these may also be applied to *B. subtilis* sRNAs and to other classes of regulatory RNAs. A great example of this is the recent termSeq approach used to identify TA systems (57). Last, it is conceivable that a systematic removal of asRNAs by genome editing or genome synthesis may become feasible in the not-too-far future. For instance, it is nowadays much easier to make specific mutations and knockdowns than a few years ago, as illustrated by a dedicated CRISPR interference approach that was recently used to create knockdowns of every essential gene in *B. subtilis* (143). Approaches of this kind could allow a functional analysis of asRNA-mediated gene regulation on the genome-wide scale. Such a challenging enterprise could specifically target the promoter regions from which asRNAs arise, as was previously explored to assess the antisense regulation of *rpsD* by S1136/S1134 (70).

In conclusion, the aim of this review was to provide a background to what is currently known about RNA-mediated regulation in the Gram-positive model organism *B. subtilis* and to provide context and considerations for future studies on the putative regulatory RNAs previously identified. Altogether, we believe that the presented structural and conservational analyses provide relevant new leads to ease the selection and in-depth analysis of the most likely functional regulatory RNAs. The studies reviewed here may thus facilitate the identification of new regulatory RNA functions in *B. subtilis*, with implications for fundamental studies on, for instance, cell differentiation and stress responses or more-applied questions in biotechnology. The new principles thus discovered can then be validated or challenged in more distantly related Gram-positive bacteria, such as *S. aureus*, as was recently exemplified by the studies by Mader et al. (33) on the roles of alternative sigma factors and transcriptional termination in the potentially spurious synthesis of asRNAs.

ACKNOWLEDGMENTS

This work was supported by the Commission of the European Union (projects LSHG-CT-2006-037469, 244093, and H2020-MSCA-ITN-2014-642836-ProteinFactory) and the transnational Systems Biology of Microorganisms (SysMO) organization (project BACELL SysMO2) through the Research Council for Earth and Life Sciences of the Netherlands Organization for Scientific Research.

REFERENCES

1. **Cech TR, Steitz JA.** 2014. The noncoding RNA revolution—trashing old rules to forge new ones. Cell **157:**77–94. http://dx.doi.org/10.1016/j.cell.2014.03.008.
2. **Waters LS, Storz G.** 2009. Regulatory RNAs in bacteria. Cell **136:**615–628. http://dx.doi.org/10.1016/j.cell.2009.01.043.
3. **Wang K, Chang H.** 2011. Molecular mechanisms of long noncoding RNAs. Mol Cell **43:**904–914. http://dx.doi.org/10.1016/j.molcel.2011.08.018.
4. **Marraffini LA, Sontheimer EJ.** 2010. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet **11:**181–190. http://dx.doi.org/10.1038/nrg2749.
5. **Schroeder R, Barta A, Semrad K.** 2004. Strategies for RNA folding and

assembly. Nat Rev Mol Cell Biol **5:**908–919. http://dx.doi.org/10.1038/nrm1497.
6. **Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY.** 2011. Understanding the transcriptome through RNA structure. Nat Rev Genet **12:**641–655. http://dx.doi.org/10.1038/nrg3049.
7. **Peer A, Margalit H.** 2011. Accessibility and evolutionary conservation mark bacterial small-RNA target-binding regions. J Bacteriol **193:**1690–1701. http://dx.doi.org/10.1128/JB.01419-10.
8. **Busch A, Richter AS, Backofen R.** 2008. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinformatics **24:**2849–2856. http://dx.doi.org/10.1093/bioinformatics/btn544.
9. **Freyhult E, Gardner PP, Moulton V.** 2005. A comparison of RNA folding measures. BMC Bioinformatics **6:**241. http://dx.doi.org/10.1186/1471-2105-6-241.
10. **Livny J, Waldor MK.** 2007. Identification of small RNAs in diverse bacterial species. Curr Opin Microbiol **10:**96–101. http://dx.doi.org/10.1016/j.mib.2007.03.005.
11. **Silvaggi JM, Perkins JB, Losick R.** 2006. Genes for small, noncoding RNAs under sporulation control in *Bacillus subtilis*. J Bacteriol **188:**532–541. http://dx.doi.org/10.1128/JB.188.2.532-541.2006.
12. **Saito S, Kakeshita H, Nakamura K.** 2009. Novel small RNA-encoding genes in the intergenic regions of *Bacillus subtilis*. Gene **428:**2–8. http://dx.doi.org/10.1016/j.gene.2008.09.024.
13. **Sharma CM, Vogel J.** 2014. Differential RNA-seq: the approach behind and the biological insight gained. Curr Opin Microbiol **19C:**97–105. http://dx.doi.org/10.1016/j.mib.2014.06.010.
14. **Irnov I, Sharma CM, Vogel J, Winkler WC.** 2010. Identification of regulatory RNAs in *Bacillus subtilis*. Nucleic Acids Res **38:**6637–6651. http://dx.doi.org/10.1093/nar/gkq454.
15. **Nicolas P, Mader U, Dervyn E, Rochat T, Leduc A, Pigeonneau N, Bidnenko E, Marchadier E, Hoebeke M, Aymerich S, Becher D, Bisicchia P, Botella E, Delumeau O, Doherty G, Denham EL, Fogg MJ, Fromion V, Goelzer A, Hansen A, Hartig E, Harwood CR, Homuth G, Jarmer H, Jules M, Klipp E, Le Chat L, Lecointe F, Lewis P, Liebermeister W, March A, Mars RA, Nannapaneni P, Noone D, Pohl S, Rinn B, Rugheimer F, Sappa PK, Samson F, Schaffer M, Schwikowski B, Steil L, Stulke J, Wiegert T, Devine KM, Wilkinson AJ, van Dijl JM, Hecker M, Volker U, Bessieres P, Noirot P.** 2012. Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. Science **335:**1103–1106. http://dx.doi.org/10.1126/science.1206848.
16. **Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J.** 2012. An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. EMBO J **31:**4005–4019. http://dx.doi.org/10.1038/emboj.2012.229.
17. **Wade JT, Grainger DC.** 2014. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. Nat Rev Microbiol **12:**647–653. http://dx.doi.org/10.1038/nrmicro3316.
18. **Earl AM, Losick R, Kolter R.** 2008. Ecology and genomics of *Bacillus subtilis*. Trends Microbiol **16:**269–275. http://dx.doi.org/10.1016/j.tim.2008.03.004.
19. **Michna RH, Commichau FM, Todter D, Zschiedrich CP, Stulke J.** 2014. SubtiWiki—a database for the model organism *Bacillus subtilis* that links pathway, interaction and expression information. Nucleic Acids Res **42:**D692–D698. http://dx.doi.org/10.1093/nar/gkt1002.
20. **Harwood CR.** 1992. *Bacillus subtilis* and its relatives: molecular biological and industrial workhorses. Trends Biotechnol **10:**247–256. http://dx.doi.org/10.1016/0167-7799(92)90233-L.
21. **van Dijl JM, Hecker M.** 2013. *Bacillus subtilis*: from soil bacterium to super-secreting cell factory. Microb Cell Fact **12:**3. http://dx.doi.org/10.1186/1475-2859-12-3.
22. **Jousselin A, Metzinger L, Felden B.** 2009. On the facultative requirement of the bacterial RNA chaperone, Hfq. Trends Microbiol **17:**399–405. http://dx.doi.org/10.1016/j.tim.2009.06.003.
23. **Hammerle H, Amman F, Vecerek B, Stulke J, Hofacker I, Blasi U.** 2014. Impact of Hfq on the *Bacillus subtilis* transcriptome. PLoS One **9:**e98661. http://dx.doi.org/10.1371/journal.pone.0098661.
24. **Mars RA, Nicolas P, Ciccolini M, Reilman E, Reder A, Schaffer M, Mader U, Volker U, van Dijl JM, Denham EL.** 2015. Small regulatory RNA-induced growth rate heterogeneity of *Bacillus subtilis*. PLoS Genet **11:**e1005046. http://dx.doi.org/10.1371/journal.pgen.1005046.
25. **Tjalsma H, Antelmann H, Jongbloed JD, Braun PG, Darmon E, Dorenbos R, Dubois JY, Westers H, Zanen G, Quax WJ, Kuipers OP, Bron S, Hecker M, van Dijl JM.** 2004. Proteomics of protein secretion by *Bacillus subtilis*: separating the "secrets" of the secretome. Microbiol Mol Biol Rev **68:**207–233. http://dx.doi.org/10.1128/MMBR.68.2.207-233.2004.
26. **Tjalsma H, Bolhuis A, Jongbloed JD, Bron S, van Dijl JM.** 2000. Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. Microbiol Mol Biol Rev **64:**515–547. http://dx.doi.org/10.1128/MMBR.64.3.515-547.2000.
27. **Brantl S, Brückner R.** 2014. Small regulatory RNAs from low-GC Gram-positive bacteria. RNA Biol **11:**443–456. http://dx.doi.org/10.4161/rna.28036.
28. **Wassarman KM.** 2007. 6S RNA: a small RNA regulator of transcription. Curr Opin Microbiol **10:**164–168. http://dx.doi.org/10.1016/j.mib.2007.03.008.
29. **Breaker RR.** 2012. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol **4:**a003566. http://dx.doi.org/10.1101/cshperspect.a003566.
30. **Hubner S, Declerck N, Diethmaier C, Le Coq D, Aymerich S, Stulke J.** 2011. Prevention of cross-talk in conserved regulatory systems: identification of specificity determinants in RNA-binding anti-termination proteins of the BglG family. Nucleic Acids Res **39:**4360–4372. http://dx.doi.org/10.1093/nar/gkr021.
31. **Gutierrez-Preciado A, Henkin TM, Grundy FJ, Yanofsky C, Merino E.** 2009. Biochemical features and functional implications of the RNA-based T-box regulatory mechanism. Microbiol Mol Biol Rev **73:**36–61. http://dx.doi.org/10.1128/MMBR.00026-08.
32. **Deiorio-Haggar K, Anthony J, Meyer MM.** 2013. RNA structures regulating ribosomal protein biosynthesis in bacilli. RNA Biol **10:**1180–1184. http://dx.doi.org/10.4161/rna.24151.
33. **Mader U, Nicolas P, Depke M, Pane-Farre J, Debarbouille M, van der Kooi-Pol MM, Guerin C, Derozier S, Hiron A, Jarmer H, Leduc A, Michalik S, Reilman E, Schaffer M, Schmidt F, Bessieres P, Noirot P, Hecker M, Msadek T, Volker U, van Dijl JM.** 2016. Staphylococcus aureus transcriptome architecture: from laboratory to infection-mimicking conditions. PLoS Genet **12:**e1005962. http://dx.doi.org/10.1371/journal.pgen.1005962.
34. **Mader U, Nicolas P, Richard H, Bessieres P, Aymerich S.** 2011. Comprehensive identification and quantification of microbial transcriptomes by genome-wide unbiased methods. Curr Opin Biotechnol **22:**32–41. http://dx.doi.org/10.1016/j.copbio.2010.10.003.
35. **Rasmussen S, Nielsen HB, Jarmer H.** 2009. The transcriptionally active regions in the genome of *Bacillus subtilis*. Mol Microbiol **73:**1043–1057. http://dx.doi.org/10.1111/j.1365-2958.2009.06830.x.
36. **Mirauta B, Nicolas P, Richard H.** 2014. Parseq: reconstruction of microbial transcription landscape from RNA-Seq read counts using state-space models. Bioinformatics **30:**1409–1416. http://dx.doi.org/10.1093/bioinformatics/btu042.
37. **Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A.** 2013. Rfam 11.0: 10 years of RNA families. Nucleic Acids Res **41:**D226–D232. http://dx.doi.org/10.1093/nar/gks1005.
38. **Abreu-Goodger C, Merino E.** 2005. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. Nucleic Acids Res **33:**W690–W692. http://dx.doi.org/10.1093/nar/gki445.
39. **Barbe V, Cruveiller S, Kunst F, Lenoble P, Meurice G, Sekowska A, Vallenet D, Wang T, Moszer I, Medigue C, Danchin A.** 2009. From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. Microbiology **155:**1758–1775. http://dx.doi.org/10.1099/mic.0.027839-0.
40. **Sierro N, Makita Y, de Hoon M, Nakai K.** 2008. DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. Nucleic Acids Res **36:**D93–D96.
41. **de Jong A, Pietersma H, Cordes M, Kuipers OP, Kok J.** 2012. PePPER: a webserver for prediction of prokaryote promoter elements and regulons. BMC Genomics **13:**299. http://dx.doi.org/10.1186/1471-2164-13-299.
42. **Campbell EA, Westblade LF, Darst SA.** 2008. Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. Curr Opin Microbiol **11:**121–127. http://dx.doi.org/10.1016/j.mib.2008.02.016.
43. **Gaballa A, Antelmann H, Aguilar C, Khakh SK, Song KB, Smaldone GT, Helmann JD.** 2008. The *Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins.

Proc Natl Acad Sci U S A **105:**11927–11932. http://dx.doi.org/10.1073/pnas.0711752105.
44. **Licht A, Preis S, Brantl S.** 2005. Implication of CcpN in the regulation of a novel untranslated RNA (SR1) in *Bacillus subtilis*. Mol Microbiol **58:**189–206. http://dx.doi.org/10.1111/j.1365-2958.2005.04810.x.
45. **Haugen SP, Ross W, Gourse RL.** 2008. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nat Rev Microbiol **6:**507–519. http://dx.doi.org/10.1038/nrmicro1912.
46. **Hofacker IL, Stadler PF.** 2006. Memory efficient folding algorithms for circular RNA secondary structures. Bioinformatics **22:**1172–1176. http://dx.doi.org/10.1093/bioinformatics/btl023.
47. **Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E.** 2010. Genome-wide measurement of RNA secondary structure in yeast. Nature **467:**103–107. http://dx.doi.org/10.1038/nature09322.
48. **Smolke CD, Keasling JD.** 2002. Effect of gene location, mRNA secondary structures, and RNase sites on expression of two genes in an engineered operon. Biotechnol Bioeng **80:**762–776. http://dx.doi.org/10.1002/bit.10434.
49. **Sharp JS, Bechhofer DH.** 2005. Effect of 5′-proximal elements on decay of a model mRNA in *Bacillus subtilis*. Mol Microbiol **57:**484–495. http://dx.doi.org/10.1111/j.1365-2958.2005.04683.x.
50. **Reilman E, Mars RA, van Dijl JM, Denham EL.** 2014. The multidrug ABC transporter BmrC/BmrD of *Bacillus subtilis* is regulated via a ribosome-mediated transcriptional attenuation mechanism. Nucleic Acids Res **42:**11393–11407. http://dx.doi.org/10.1093/nar/gku832.
51. **Lehnik-Habrink M, Pfortner H, Rempeters L, Pietack N, Herzberg C, Stulke J.** 2010. The RNA degradosome in *Bacillus subtilis*: identification of CshA as the major RNA helicase in the multiprotein complex. Mol Microbiol **77:**958–971. http://dx.doi.org/10.1111/j.1365-2958.2010.07264.x.
52. **Lehnik-Habrink M, Lewis RJ, Mader U, Stulke J.** 2012. RNA degradation in *Bacillus subtilis*: an interplay of essential endo- and exoribonucleases. Mol Microbiol **84:**1005–1017. http://dx.doi.org/10.1111/j.1365-2958.2012.08072.x.
53. **Durand S, Gilet L, Bessieres P, Nicolas P, Condon C.** 2012. Three essential ribonucleases—RNase Y, J1, and III—control the abundance of a majority of *Bacillus subtilis* mRNAs. PLoS Genet **8:**e1002520. http://dx.doi.org/10.1371/journal.pgen.1002520.
54. **Mathy N, Benard L, Pellegrini O, Daou R, Wen T, Condon C.** 2007. 5′-to-3′ exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5′ stability of mRNA. Cell **129:**681–692. http://dx.doi.org/10.1016/j.cell.2007.02.051.
55. **Durand S, Tomasini A, Braun F, Condon C, Romby P.** 2015. sRNA and mRNA turnover in Gram-positive bacteria. FEMS Microbiol Rev **39:**316–330. http://dx.doi.org/10.1093/femsre/fuv007.
56. **Naville M, Gautheret D.** 2010. Transcription attenuation in bacteria: theme and variations. Brief Funct Genomics **9:**178–189. http://dx.doi.org/10.1093/bfgp/elq008.
57. **Dar D, Shamir M, Mellin JR, Koutero M, Stern-Ginossar N, Cossart P, Sorek R.** 2016. Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. Science **352:**aad9822. http://dx.doi.org/10.1126/science.aad9822.
58. **Ohki R, Tateno K, Takizawa T, Aiso T, Murata M.** 2005. Transcriptional termination control of a novel ABC transporter gene involved in antibiotic resistance in *Bacillus subtilis*. J Bacteriol **187:**5946–5954. http://dx.doi.org/10.1128/JB.187.17.5946-5954.2005.
59. **Saito H, Inoue T.** 2009. Synthetic biology with RNA motifs. Int J Biochem Cell Biol **41:**398–404. http://dx.doi.org/10.1016/j.biocel.2008.08.017.
60. **Qi LS, Arkin AP.** 2014. A versatile framework for microbial engineering using synthetic non-coding RNAs. Nat Rev Microbiol **12:**341–354. http://dx.doi.org/10.1038/nrmicro3244.
61. **Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR.** 2003. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. Cell **113:**577–586. http://dx.doi.org/10.1016/S0092-8674(03)00391-X.
62. **Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, Lee M, Roth A, Sudarsan N, Jona I, Wickiser JK, Breaker RR.** 2004. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. Proc Natl Acad Sci U S A **101:**6421–6426. http://dx.doi.org/10.1073/pnas.0308014101.
63. **Mehne FM, Gunka K, Eilers H, Herzberg C, Kaever V, Stulke J.** 2013. Cyclic di-AMP homeostasis in *Bacillus subtilis*: both lack and high level accumulation of the nucleotide are detrimental for cell growth. J Biol Chem **288:**2004–2017. http://dx.doi.org/10.1074/jbc.M112.395491.
64. **Nelson JW, Sudarsan N, Furukawa K, Weinberg Z, Wang JX, Breaker RR.** 2013. Riboswitches in eubacteria sense the second messenger c-di-AMP. Nat Chem Biol **9:**834–839. http://dx.doi.org/10.1038/nchembio.1363.
65. **Winkler WC, Cohen-Chalamish S, Breaker RR.** 2002. An mRNA structure that controls gene expression by binding FMN. Proc Natl Acad Sci U S A **99:**15908–15913. http://dx.doi.org/10.1073/pnas.212628899.
66. **Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, Breaker RR.** 2010. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. Genome Biol **11:**R31. http://dx.doi.org/10.1186/gb-2010-11-3-r31.
67. **Henkin TM, Glass BL, Grundy FJ.** 1992. Analysis of the *Bacillus subtilis* tyrS gene: conservation of a regulatory sequence in multiple tRNA synthetase genes. J Bacteriol **174:**1299–1306.
68. **Yakhnin H, Yakhnin AV, Babitzke P.** 2015. Ribosomal protein L10(L12)4 autoregulates expression of the *Bacillus subtilis* rplJL operon by a transcription attenuation mechanism. Nucleic Acids Res **43:**7032–7043. http://dx.doi.org/10.1093/nar/gkv628.
69. **Shajani Z, Sykes MT, Williamson JR.** 2011. Assembly of bacterial ribosomes. Annu Rev Biochem **80:**501–526. http://dx.doi.org/10.1146/annurev-biochem-062608-160432.
70. **Mars RA, Mendonca K, Denham EL, van Dijl JM.** 2015. The reduction in small ribosomal subunit abundance in ethanol-stressed cells of *Bacillus subtilis* is mediated by a SigB-dependent antisense RNA. Biochim Biophys Acta **1853:**2553–2559. http://dx.doi.org/10.1016/j.bbamcr.2015.06.009.
71. **Babitzke P.** 2004. Regulation of transcription attenuation and translation initiation by allosteric control of an RNA-binding protein: the *Bacillus subtilis* TRAP protein. Curr Opin Microbiol **7:**132–139. http://dx.doi.org/10.1016/j.mib.2004.02.003.
72. **Willimsky G, Bang H, Fischer G, Marahiel MA.** 1992. Characterization of cspB, a *Bacillus subtilis* inducible cold shock gene affecting cell viability at low temperatures. J Bacteriol **174:**6326–6335.
73. **Graumann P, Wendrich TM, Weber MH, Schroder K, Marahiel MA.** 1997. A family of cold shock proteins in *Bacillus subtilis* is essential for cellular growth and for efficient protein synthesis at optimal and low temperatures. Mol Microbiol **25:**741–756. http://dx.doi.org/10.1046/j.1365-2958.1997.5121878.x.
74. **Turner RJ, Lu Y, Switzer RL.** 1994. Regulation of the *Bacillus subtilis* pyrimidine biosynthetic (pyr) gene cluster by an autogenous transcriptional attenuation mechanism. J Bacteriol **176:**3708–3722.
75. **Lu Y, Turner RJ, Switzer RL.** 1996. Function of RNA secondary structures in transcriptional attenuation of the *Bacillus subtilis* pyr operon. Proc Natl Acad Sci U S A **93:**14462–14467. http://dx.doi.org/10.1073/pnas.93.25.14462.
76. **Glatz E, Nilsson RP, Rutberg L, Rutberg B.** 1996. A dual role for the *Bacillus subtilis* glpD leader and the GlpP protein in the regulated expression of glpD: antitermination and control of mRNA stability. Mol Microbiol **19:**319–328. http://dx.doi.org/10.1046/j.1365-2958.1996.376903.x.
77. **Kumarevel T, Mizuno H, Kumar PK.** 2005. Structural basis of HutP-mediated anti-termination and roles of the $Mg2+$ ion and L-histidine ligand. Nature **434:**183–191. http://dx.doi.org/10.1038/nature03355.
78. **Morinaga T, Kobayashi K, Ashida H, Fujita Y, Yoshida K.** 2010. Transcriptional regulation of the *Bacillus subtilis* asnH operon and role of the 5′-proximal long sequence triplication in RNA stabilization. Microbiology **156:**1632–1641. http://dx.doi.org/10.1099/mic.0.036582-0.
79. **Hambraeus G, von Wachenfeldt C, Hederstedt L.** 2003. Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. Mol Genet Genomics **269:**706–714. http://dx.doi.org/10.1007/s00438-003-0883-6.
80. **Noone D, Salzberg LI, Botella E, Basell K, Becher D, Antelmann H, Devine KM.** 2014. A highly unstable transcript makes CwlO D,L-endopeptidase expression responsive to growth conditions in *Bacillus subtilis*. J Bacteriol **196:**237–247. http://dx.doi.org/10.1128/JB.00986-13.
81. **Deana A, Belasco JG.** 2005. Lost in translation: the influence of ribosomes on bacterial mRNA decay. Genes Dev **19:**2526–2533. http://dx.doi.org/10.1101/gad.1348805.
82. **Yao S, Bechhofer DH.** 2009. Processing and stability of inducibly expressed rpsO mRNA derivatives in *Bacillus subtilis*. J Bacteriol **191:**5680–5689. http://dx.doi.org/10.1128/JB.00740-09.

83. **Hambraeus G, Karhumaa K, Rutberg B.** 2002. A 5′ stem-loop and ribosome binding but not translation are important for the stability of *Bacillus subtilis* aprE leader mRNA. Microbiology **148:**1795–1803. http://dx.doi.org/10.1099/00221287-148-6-1795.
84. **de Smit MH, van Duin J.** 1994. Control of translation by mRNA secondary structure in Escherichia coli. A quantitative analysis of literature data. J Mol Biol **244:**144–150.
85. **Kruger S, Gertz S, Hecker M.** 1996. Transcriptional analysis of bglPH expression in *Bacillus subtilis*: evidence for two distinct pathways mediating carbon catabolite repression. J Bacteriol **178:**2637–2644.
86. **Meng Q, Switzer RL.** 2001. Regulation of transcription of the *Bacillus subtilis* pyrG gene, encoding cytidine triphosphate synthetase. J Bacteriol **183:**5513–5522. http://dx.doi.org/10.1128/JB.183.19.5513-5522.2001.
87. **Jensen-MacAllister IE, Meng Q, Switzer RL.** 2007. Regulation of pyrG expression in *Bacillus subtilis*: CTP-regulated antitermination and reiterative transcription with pyrG templates in vitro. Mol Microbiol **63:**1440–1452. http://dx.doi.org/10.1111/j.1365-2958.2007.05595.x.
88. **Lehnik-Habrink M, Schaffer M, Mader U, Diethmaier C, Herzberg C, Stulke J.** 2011. RNA processing in *Bacillus subtilis*: identification of targets of the essential RNase Y. Mol Microbiol **81:**1459–1473. http://dx.doi.org/10.1111/j.1365-2958.2011.07777.x.
89. **Nahvi A, Barrick JE, Breaker RR.** 2004. Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. Nucleic Acids Res **32:**143–150. http://dx.doi.org/10.1093/nar/gkh167.
90. **Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM.** 2005. Animal microRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. Cell **123:**1133–1146. http://dx.doi.org/10.1016/j.cell.2005.11.023.
91. **Tomecki R, Dziembowski A.** 2010. Novel endoribonucleases as central players in various pathways of eukaryotic RNA metabolism. RNA **16:**1692–1724. http://dx.doi.org/10.1261/rna.2237610.
92. **Belasco JG.** 2010. All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. Nat Rev Mol Cell Biol **11:**467–478. http://dx.doi.org/10.1038/nrm2917.
93. **Gerwig J, Stulke J.** 2014. Caught in the act: RNA-Seq provides novel insights into mRNA degradation. Mol Microbiol **94:**5–8. http://dx.doi.org/10.1111/mmi.12769.
94. **Alen C, Sonenshein AL.** 1999. *Bacillus subtilis* aconitase is an RNA-binding protein. Proc Natl Acad Sci U S A **96:**10412–10417. http://dx.doi.org/10.1073/pnas.96.18.10412.
95. **Miethke M, Monteferrante CG, Marahiel MA, van Dijl JM.** 2013. The *Bacillus subtilis* EfeUOB transporter is essential for high-affinity acquisition of ferrous and ferric iron. Biochim Biophys Acta **1833:**2267–2278. http://dx.doi.org/10.1016/j.bbamcr.2013.05.027.
96. **Pfleger BF, Pitera DJ, Smolke CD, Keasling JD.** 2006. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol **24:**1027–1032. http://dx.doi.org/10.1038/nbt1226.
97. **Allenby NE, O'Connor N, Pragai Z, Carter NM, Miethke M, Engelmann S, Hecker M, Wipat A, Ward AC, Harwood CR.** 2004. Post-transcriptional regulation of the *Bacillus subtilis* pst operon encoding a phosphate-specific ABC transporter. Microbiology **150:**2619–2628. http://dx.doi.org/10.1099/mic.0.27126-0.
98. **Irnov I, Winkler WC.** 2010. A regulatory RNA required for antitermination of biofilm and capsular polysaccharide operons in Bacillales. Mol Microbiol **76:**559–575. http://dx.doi.org/10.1111/j.1365-2958.2010.07131.x.
99. **Jerga A, Lu YJ, Schujman GE, de Mendoza D, Rock CO.** 2007. Identification of a soluble diacylglycerol kinase required for lipoteichoic acid production in *Bacillus subtilis*. J Biol Chem **282:**21738–21745. http://dx.doi.org/10.1074/jbc.M703536200.
100. **Smaldone GT, Antelmann H, Gaballa A, Helmann JD.** 2012. The FsrA sRNA and FbpB protein mediate the iron-dependent induction of the *Bacillus subtilis* lutABC iron-sulfur-containing oxidases. J Bacteriol **194:**2586–2593. http://dx.doi.org/10.1128/JB.05567-11.
101. **Updegrove TB, Shabalina SA, Storz G.** 2015. How do base-pairing small RNAs evolve? FEMS Microbiol Rev **39:**379–391. http://dx.doi.org/10.1093/femsre/fuv014.
102. **Smaldone GT, Revelles O, Gaballa A, Sauer U, Antelmann H, Helmann JD.** 2012. A global investigation of the *Bacillus subtilis* iron-sparing response identifies major changes in metabolism. J Bacteriol **194:**2594–2605. http://dx.doi.org/10.1128/JB.05990-11.
103. **Beisel CL, Storz G.** 2010. Base pairing small RNAs and their roles in global regulatory networks. FEMS Microbiol Rev **34:**866–882. http://dx.doi.org/10.1111/j.1574-6976.2010.00241.x.
104. **Heidrich N, Chinali A, Gerth U, Brantl S.** 2006. The small untranslated RNA SR1 from the *Bacillus subtilis* genome is involved in the regulation of arginine catabolism. Mol Microbiol **62:**520–536. http://dx.doi.org/10.1111/j.1365-2958.2006.05384.x.
105. **Gimpel M, Heidrich N, Mader U, Krugel H, Brantl S.** 2010. A dual-function sRNA from B. subtilis: SR1 acts as a peptide encoding mRNA on the gapA operon. Mol Microbiol **76:**990–1009. http://dx.doi.org/10.1111/j.1365-2958.2010.07158.x.
106. **Gimpel M, Preis H, Barth E, Gramzow L, Brantl S.** 2012. SR1—a small RNA with two remarkably conserved functions. Nucleic Acids Res **40:**11659–11672. http://dx.doi.org/10.1093/nar/gks895.
107. **Geissmann T, Chevalier C, Cros MJ, Boisset S, Fechter P, Noirot C, Schrenzel J, Francois P, Vandenesch F, Gaspin C, Romby P.** 2009. A search for small noncoding RNAs in Staphylococcus aureus reveals a conserved sequence motif for regulation. Nucleic Acids Res **37:**7239–7257. http://dx.doi.org/10.1093/nar/gkp668.
108. **Durand S, Braun F, Lioliou E, Romilly C, Helfer AC, Kuhn L, Quittot N, Nicolas P, Romby P, Condon C.** 2015. A nitric oxide regulated small RNA controls expression of genes involved in redox homeostasis in *Bacillus subtilis*. PLoS Genet **11:**e1004957. http://dx.doi.org/10.1371/journal.pgen.1004957.
109. **Schmalisch M, Maiques E, Nikolov L, Camp AH, Chevreux B, Muffler A, Rodriguez S, Perkins J, Losick R.** 2010. Small genes under sporulation control in the *Bacillus subtilis* genome. J Bacteriol **192:**5402–5412. http://dx.doi.org/10.1128/JB.00534-10.
110. **Silvaggi JM, Perkins JB, Losick R.** 2005. Small untranslated RNA antitoxin in *Bacillus subtilis*. J Bacteriol **187:**6641–6650. http://dx.doi.org/10.1128/JB.187.19.6641-6650.2005.
111. **Commichau FM, Stulke J.** 2012. A mystery unraveled: essentiality of RNase III in *Bacillus subtilis* is caused by resident prophages. PLoS Genet **8:**e1003199. http://dx.doi.org/10.1371/journal.pgen.1003199.
112. **Durand S, Jahn N, Condon C, Brantl S.** 2012. Type I toxin-antitoxin systems in *Bacillus subtilis*. RNA Biol **9:**1491–1497. http://dx.doi.org/10.4161/rna.22358.
113. **Jahn N, Preis H, Wiedemann C, Brantl S.** 2012. BsrG/SR4 from *Bacillus subtilis*—the first temperature-dependent type I toxin-antitoxin system. Mol Microbiol **83:**579–598. http://dx.doi.org/10.1111/j.1365-2958.2011.07952.x.
114. **Jahn N, Brantl S.** 2013. One antitoxin—two functions: SR4 controls toxin mRNA decay and translation. Nucleic Acids Res **41:**9870–9880. http://dx.doi.org/10.1093/nar/gkt735.
115. **Meissner C, Jahn N, Brantl S.** 2016. In vitro characterization of the type I toxin-antitoxin system bsrE/SR5 from *Bacillus subtilis*. J Biol Chem **291:**560–571. http://dx.doi.org/10.1074/jbc.M115.697524.
116. **Loh E, Dussurget O, Gripenland J, Vaitkevicius K, Tiensuu T, Mandin P, Repoila F, Buchrieser C, Cossart P, Johansson J.** 2009. A trans-acting riboswitch controls expression of the virulence regulator PrfA in Listeria monocytogenes. Cell **139:**770–779. http://dx.doi.org/10.1016/j.cell.2009.08.046.
117. **Marchais A, Duperrier S, Durand S, Gautheret D, Stragier P.** 2011. CsfG, a sporulation-specific, small non-coding RNA highly conserved in endospore formers. RNA Biol **8:**358–364. http://dx.doi.org/10.4161/rna.8.3.14998.
118. **Georg J, Hess WR.** 2011. *cis*-Antisense RNA, another level of gene regulation in bacteria. Microbiol Mol Biol Rev **75:**286–300. http://dx.doi.org/10.1128/MMBR.00032-10.
119. **Brantl S.** 2007. Regulatory mechanisms employed by cis-encoded antisense RNAs. Curr Opin Microbiol **10:**102–109. http://dx.doi.org/10.1016/j.mib.2007.03.012.
120. **Thomason MK, Storz G.** 2010. Bacterial antisense RNAs: how many are there, and what are they doing? Annu Rev Genet **44:**167–188. http://dx.doi.org/10.1146/annurev-genet-102209-163523.
121. **Eiamphungporn W, Helmann JD.** 2009. Extracytoplasmic function sigma factors regulate expression of the *Bacillus subtilis* yabE gene via a *cis*-acting antisense RNA. J Bacteriol **191:**1101–1105. http://dx.doi.org/10.1128/JB.01530-08.
122. **Luo Y, Helmann JD.** 2012. A sigmaD-dependent antisense transcript modulates expression of the cyclic-di-AMP hydrolase GdpP in *Bacillus subtilis*. Microbiology **158:**2732–2741. http://dx.doi.org/10.1099/mic.0.062174-0.
123. **Guell M, van Noort V, Yus E, Chen WH, Leigh-Bell J, Michalodimi-**

trakis K, Yamada T, Arumugam M, Doerks T, Kuhner S, Rode M, Suyama M, Schmidt S, Gavin AC, Bork P, Serrano L. 2009. Transcriptome complexity in a genome-reduced bacterium. Science **326:**1268–1271. http://dx.doi.org/10.1126/science.1176951.

124. **Hecker M, Pane-Farre J, Volker U.** 2007. SigB-dependent general stress response in *Bacillus subtilis* and related gram-positive bacteria. Annu Rev Microbiol **61:**215–236. http://dx.doi.org/10.1146/annurev.micro.61.080706.093445.
125. **Weng M, Nagy PL, Zalkin H.** 1995. Identification of the *Bacillus subtilis* pur operon repressor. Proc Natl Acad Sci U S A **92:**7455–7459. http://dx.doi.org/10.1073/pnas.92.16.7455.
126. **Freese E, Heinze JE, Galliers EM.** 1979. Partial purine deprivation causes sporulation of *Bacillus subtilis* in the presence of excess ammonia, glucose and phosphate. J Gen Microbiol **115:**193–205. http://dx.doi.org/10.1099/00221287-115-1-193.
127. **Eijlander RT, de Jong A, Krawczyk AO, Holsappel S, Kuipers OP.** 2014. SporeWeb: an interactive journey through the complete sporulation cycle of *Bacillus subtilis*. Nucleic Acids Res **42:**D685–D691. http://dx.doi.org/10.1093/nar/gkt1007.
128. **Kobayashi K, Shoji K, Shimizu T, Nakano K, Sato T, Kobayashi Y.** 1995. Analysis of a suppressor mutation ssb (kinC) of sur0B20 (spo0A) mutation in *Bacillus subtilis* reveals that kinC encodes a histidine protein kinase. J Bacteriol **177:**176–182.
129. **Ingram LO, Vreeland NS.** 1980. Differential effects of ethanol and hexanol on the Escherichia coli cell envelope. J Bacteriol **144:**481–488.
130. **Reichmann NT, Grundling A.** 2011. Location, synthesis and function of glycolipids and polyglycerolphosphate lipoteichoic acid in Gram-positive bacteria of the phylum Firmicutes. FEMS Microbiol Lett **319:**97–105. http://dx.doi.org/10.1111/j.1574-6968.2011.02260.x.
131. **Vogel J, Luisi BF.** 2011. Hfq and its constellation of RNA. Nat Rev Microbiol **9:**578–589. http://dx.doi.org/10.1038/nrmicro2615.
132. **Rochat T, Delumeau O, Figueroa-Bossi N, Noirot P, Bossi L, Dervyn E, Bouloc P.** 2015. Tracking the elusive function of *Bacillus subtilis* Hfq. PLoS One **10:**e0124977. http://dx.doi.org/10.1371/journal.pone.0124977.
133. **Boisset S, Geissmann T, Huntzinger E, Fechter P, Bendridi N, Possedko M, Chevalier C, Helfer AC, Benito Y, Jacquier A, Gaspin C, Vandenesch F, Romby P.** 2007. Staphylococcus aureus RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. Genes Dev **21:**1353–1366. http://dx.doi.org/10.1101/gad.423507.
134. **Burke TP, Portnoy DA.** 2016. SpoVG is a conserved RNA-binding protein that regulates Listeria monocytogenes lysozyme resistance, virulence, and swarming motility. mBio **7:**e00240. http://dx.doi.org/10.1128/mBio.00240-16.
135. **Vrentas C, Ghirlando R, Keefer A, Hu Z, Tomczak A, Gittis AG, Murthi A, Garboczi DN, Gottesman S, Leppla SH.** 2015. Hfqs in Bacillus anthracis: role of protein sequence variation in the structure and function of proteins in the Hfq family. Protein Sci **24:**1808–1819. http://dx.doi.org/10.1002/pro.2773.
136. **Tanay A, Regev A, Shamir R.** 2005. Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. Proc Natl Acad Sci U S A **102:**7203–7208. http://dx.doi.org/10.1073/pnas.0502521102.
137. **Goelzer A, Fromion V.** 2011. Bacterial growth rate reflects a bottleneck in resource allocation. Biochim Biophys Acta **1810:**978–988. http://dx.doi.org/10.1016/j.bbagen.2011.05.014.
138. **Kirschner M, Gerhart J.** 1998. Evolvability. Proc Natl Acad Sci U S A **95:**8420–8427. http://dx.doi.org/10.1073/pnas.95.15.8420.
139. **Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, Charloteaux B, Hidalgo CA, Barbette J, Santhanam B, Brar GA, Weissman JS, Regev A, Thierry-Mieg N, Cusick ME, Vidal M.** 2012. Proto-genes and de novo gene birth. Nature **487:**370–374. http://dx.doi.org/10.1038/nature11184.
140. **Shimoni Y, Friedlander G, Hetzroni G, Niv G, Altuvia S, Biham O, Margalit H.** 2007. Regulation of gene expression by small non-coding RNAs: a quantitative view. Mol Syst Biol. **3:**138. http://dx.doi.org/10.1038/msb4100181.
141. **Arrieta-Ortiz ML, Hafemeister C, Bate AR, Chu T, Greenfield A, Shuster B, Barry SN, Gallitto M, Liu B, Kacmarczyk T, Santoriello F, Chen J, Rodrigues CD, Sato T, Rudner DZ, Driks A, Bonneau R, Eichenberger P.** 2015. An experimentally supported model of the *Bacillus subtilis* global transcriptional regulatory network. Mol Syst Biol **11:**839. http://dx.doi.org/10.15252/msb.20156236.
142. **Barquist L, Vogel J.** 2015. Accelerating discovery and functional analysis of small RNAs with new technologies. Annu Rev Genet **49:**367–394. http://dx.doi.org/10.1146/annurev-genet-112414-054804.
143. **Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, Hawkins JS, Lu CH, Koo BM, Marta E, Shiver AL, Whitehead EH, Weissman JS, Brown ED, Qi LS, Huang KC, Gross CA.** 2016. A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. Cell **165:**1493–1506. http://dx.doi.org/10.1016/j.cell.2016.05.003.

Ruben A. T. Mars is motivated by the challenge to move away from a descriptive understanding toward a functional understanding of genes, metabolites, and microorganisms. In 2010, he obtained his master's degree in molecular biology and biotechnology from the University of Groningen, the Netherlands. During his Ph.D. research in the laboratory of Jan Maarten van Dijl at the Department of Medical Microbiology of the University of Groningen and the University Medical Center Groningen, he worked on identifying new regulatory RNA functions in the Gram-positive bacterium *Bacillus subtilis*. After he obtained his Ph.D. in December 2014, he moved to the laboratory of Uwe Sauer at ETH Zurich, where he is applying advanced metabolomics technologies to unravel microbial interactions in ecosystems. His broader interest is to utilize microbial diversity to improve the quality of life.

Pierre Nicolas develops and uses computational methods to address challenges in interpreting omics data sets. He has a longstanding interest in the genome and transcriptome analysis of *Bacillus subtilis* and its relatives, on which he started working during his Ph.D. obtained in 2003 at the Université d'Evry after obtaining an M.Sc. degree in biomathematics at the Université Paris 7. A postdoctoral fellowship in the computational biology department of the University of Southern California gave him the opportunity to become familiar with mathematical approaches used in evolution and population genetics. His current research themes at INRA Jouy-en-Josas include identifying the border between functional transcription and noise from evolutionary and molecular sources in bacterial transcriptomes. He works also on statistical models and algorithms for *de novo* inference of bacterial regulatory networks from transcriptome data accounting for a diversity of lifestyles rather than for targeted genetic mutations.

Emma L. Denham is Assistant Professor of Molecular Bacteriology at the University of Warwick, United Kingdom. She earned her undergraduate degree in genetics from the University of Leicester, before going on and completing her Ph.D. at the Institute for Animal Health, studying protein secretion in streptococci. Dr. Denham was a postdoctoral scientist in J. M. van Dijl's lab from 2007 to 2013, where she studied gene expression and regulatory events in *Bacillus subtilis*. In 2013, she moved to the University of Warwick to set up her laboratory, where she studies the role of regulatory RNAs in the lifestyles of Gram-positive bacteria.

Jan Maarten van Dijl is Professor of Molecular Bacteriology at the University of Groningen and the University Medical Center Groningen, the Netherlands. He earned his Ph.D. in mathematics and natural sciences in 1990 under the supervision of Gerard Venema at the University of Groningen. As a postdoctoral researcher, he joined the group of Gottfried Schatz at the Biozentrum of the University of Basel, Switzerland, in the period between 1994 and 1996. Dr. van Dijl became Assistant Professor in Pharmaceutical Biology in 1998 and has been Professor of Molecular Bacteriology since 2004. His research addresses fundamental and applied aspects of gene expression, protein secretion, and antimicrobial drug resistance with a special focus on Gram-positive bacteria. He has been coordinator of various national and European research programs on protein production systems and the development of novel antimicrobial therapies.